THIS WEEK

EDITORIALS

CLIMATE Pessimists fight over a pragmatic green approach **p.444**

WORLD VIEW Rural students in China face uphill struggle **p.445**



MIGRATION The magnetic pull on monarch butterflies **p.446**

Biosafety in the balance

An accident with anthrax demonstrates that pathogen research always carries a risk of release — and highlights the need for rigorous scrutiny of gain-of-function flu studies.

The news last week of an accident involving live anthrax bacteria at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, is troubling. Some 84 workers were potentially exposed to the deadly Ames strain at three CDC labs. But the incident will cause much wider ripples: it highlights the risks of the current proliferation of biocontainment labs and work on dangerous pathogens. If an accident can happen at the CDC, then it can happen anywhere.

Details are sparse, but it seems that the anthrax was being inactivated in a biosafety-level-3 (BSL-3) high-containment lab so that it could be studied at the three BSL-2 labs. But live bacteria survived the inactivation step, and were not detected before samples were sent out. The CDC considers the risk that the exposed workers have been infected to be low, and all have been offered protective antibiotics.

Such lab accidents are fortunately not commonplace. A CDC analysis in 2012 reported, for example, that there were 727 incidents of theft, loss or release of Select Agents and Toxins in the United States between 2004 and 2010, resulting in 11 laboratory-acquired infections and no secondary transmission (R. D. Henkel *et al. Appl. Biosafety* **17,** 171–180; 2012). Anthrax is contracted by direct exposure to spores, and does not spread between people. Much more potentially dangerous are lab accidents involving agents that do. It is impossible to read about the CDC incident and not breathe a large sigh of relief that it did not involve a novel engineered pandemic influenza strain.

Groups led by Ron Fouchier of the Erasmus Medical Center in Rotterdam, the Netherlands, and Yoshihiro Kawaoka of the University of Wisconsin–Madison created a storm in late 2011 when they artificially engineered potentially pandemic forms of the H5N1 avian flu virus. In January last year, researchers ended a voluntary 12-month moratorium on such gain-of-function flu research, which can increase the host range, transmissibility or virulence of viruses (see *Nature* **493,** 460; 2013), and work resumed.

This month, Kawaoka's group reported that it had engineered a *de novo* flu virus from wild-avian-flu-strain genes that coded for proteins similar to those in the 1918 pandemic virus (T. Watanabe *Cell Host Microbe* **15,** 692–705; 2014). The researchers were able to make a virulent version that could transmit between ferrets, and they concluded that a 1918-like virus could therefore emerge from wild avian flu viruses.

In the century since the 1918 flu hit, no similar pandemic variant has emerged despite wild animal flu viruses mutating and reassorting incessantly. The 1918 H1N1 virus was reconstructed in 2005, but human immunity to it became widespread following the 2009 H1N1 pandemic. There are no mammalian-transmissible 1918-like avian flus in the wild; the only ones that exist are Kawaoka's team's engineered strains.

Researchers such as Kawaoka and Fouchier argue that by engineering mutant viruses in the lab, they can identify mutations and traits that allow the pathogens to spread between mammals. This in turn, they argue, allows assessment of the pandemic potential of animal-flu viruses. In the long term, such experiments could help to elucidate the mechanisms of virus transmissibility and pathogenicity. But their shorter-term public-health benefits have been overstated. The risks and benefits must therefore be carefully weighed, and rigorous oversight is needed to ensure that such work is done only at facilities with the highest standards of biosafety.

Other scientists argue that the concept of predicting the pandemic potential of flu viruses from mutations, although appealing, is simplistic. They say that the identified mutations are but a handful out of millions of possible combinations, many of which might also allow mammalian transmission. They argue that mutations in specific proteins cannot reliably predict traits, and that outcomes depend on interactions between various other background genetic changes throughout the virus.

"The idea of an accidental release of a potentially pandemic flu virus cannot be completely written off."

These points were highlighted in a paper in *PLoS Medicine* last month (M. Lipsitch and A. P. Galvani *PLoS Med.* **11,** e1001646; 2014), and in a letter by 56 leading virologists, infectious-disease specialists and public-health experts to European Commission president José Manuel Barroso last December (see *Nature* http://doi.org/tdb; 2013). They also question the claimed public-health benefits of such research, and argue that similar information could be obtained through safer experiments. Opponents of gain-of-function flu research call, in particular, for more rigorous risk–benefit assessments. The CDC accident shows that, should such research proliferate, the idea of an accidental release of a potentially pandemic flu virus cannot be completely written off. This demands that such research proposals receive the utmost scrutiny.

A US Government Accountability Office report released in February last year expressed concern that the proliferation of US high-containment labs following the terrorist attacks of 11 September 2011 and the anthrax-letter attacks the same year was proceeding without a rigorous assessment of the nation's real needs across all government agencies, universities and private companies. "Increasing the number of laboratories also increases the aggregate national risk," it noted. No one keeps track, for example, of how many BSL-3 labs there are in the United States alone, although their number is thought to be in the thousands. The number of such labs is increasing in China and elsewhere.

After smallpox was eradicated in 1980, there was a concerted international effort to reduce the number of labs holding stocks to just two: one at the CDC and one at the Russian State Research Center of Virology and Biotechnology in Koltsovo. All research at these centres must be approved by the World Health Organization. The fewer the labs that perform experiments, the smaller is the risk of an accidental release. But as the CDC accident reminds us, should gain-of-function flu research proliferate, in particular at facilities with less than exemplary biosafety standards, the risks of an accidental release of a potentially pandemic flu virus will be multiplied. ■

© 2014 Macmillan Publishers Limited. All rights reserved

Metrics market

Measures of research impact are improving, but universities should be wary of their limits.

With the FIFA World Cup well under way in Brazil (and certain teams already on their way home) there is much analysis of what went wrong for some and what is going right for others. In a parallel effort, marketing departments across the globe are engaged in a final push to link the events on the field with their brands and products.

The result can be a curious, even surreal, blend. Witness the announcement by analytics company Thomson Reuters, for instance, that it was kicking off its own World Cup — of research performance. In the first round, the firm announced that England could have reversed its disappointing loss against Italy had it been playing on the basis of research citation impact. By the same comparison, Australia would have defeated Chile — but the Netherlands would still have crushed Spain. Four more rounds of elimination will pit countries against each other in terms of their proportion of international collaborations, highly cited papers and relative world impact. Who can wait for the United States–Switzerland semi-final?

This gimmicky tie-in illustrates a trend that deserves attention: the retooling of the bibliometrics industry. Thomson Reuters was promoting the updated range of bibliometric indicators in its research analytics service, InCites. And it is not the only company to have refreshed its bibliometrics offerings for 2014.

Since January, Elsevier has been promoting the next generation of its SciVal product, and on 12 June, Altmetric (supported by Macmillan Science and Education, which owns Nature Publishing Group, the publishers of *Nature*) launched its own commercial offering for research institutions: a tool to track the online impact of faculty members' academic papers. One intriguing metric that was launched in April for the Lens database, run by the non-profit organization Cambia in Canberra, allows researchers to freely examine how many patents have cited their papers, although currently only the life sciences are included.

Products offered by commercial analytics firms are worth watching because they generally determine how research institutions track and assess their scholars. The latest products mean that it is now easier than ever to calculate a dizzying range of metrics for any group of papers — including aggregations of papers at the level of the individual or the faculty, as well as the country.

On the plus side, this new generation of products is more sophisticated and takes into account the insights and criticisms of bibliometrics experts. The tools now tend to focus on individual research papers as a core unit of output, rather than the journal in which a paper is published. And they increasingly recognize that it is only meaningful to compare metrics in context: for instance, normalizing for performance relative to papers of a similar age, research field and publication venue.

"It is easier than ever to calculate a dizzying range of metrics for any group of papers."

But on the minus side, it is still easy to misuse these offerings. The latest breed of bibliometrics is useful as a marketing tool for individuals, as a way to spot unnoticed pockets of excellence and, yes, even to rank countries or institutions in a research World Cup. But there is a risk that universities are buying increasingly sophisticated products to track their performance without really understanding the limitations of such metrics. On page 470, Jonathan Adams reviews *Beyond Bibliometrics*, a book that outlines the history and future direction of attempts to measure scholarly impact. The editors, Blaise Cronin and Cassidy Sugimoto, urge caution. Adams asks: "Even after decades of use, do we really understand what citation data are and what we do with them? Do those who use bibliometrics have clear criteria for how they employ and interpret them?"

As this journal has noted before (see www.nature.com/metrics), many researchers feel under stress from their attempts to maximize the metrics by which their academic success is judged. It has never been more important to demand clarity and transparency from research managers on exactly which metrics they are using to evaluate scholars, and why. With this next generation of commercial tools, it does not seem so fanciful to picture a senior manager somewhere in your research institution engaged in a kind of perpetual departmental World Cup, scrolling through screeds of bar charts and playing 'fantasy faculty' with the careers of researchers. ■

Storm warning

Environmentalists are divided over whether it is possible to have a 'good' Anthropocene.

Winston Churchill called it the enemy within. Leonardo da Vinci noted that: "As every divided kingdom falls, so every mind divided between many studies confounds and saps itself." If there is one word that sums up current attitudes to the climate problem and society's response to it, then it is 'divided'. And that describes just the attitudes of those who agree that there is a serious problem to deal with.

Those who do not accept this reality — for whatever reasons — must have been rubbing their hands this week as arguments raged among their opponents. There is a new term in town — 'ecopragmatist'. It can be used both as a compliment and an insult.

The lightning rod for the latest storm was a fairly benign talk given in New York this month by *New York Times* environment writer Andrew Revkin on the difficulty of applying numerical targets and goals, on carbon emissions say, to real-world behaviour. It might have gone largely unnoticed, except that Revkin included in the talk's title the idea of a 'good' Anthropocene — the informal name for the period, beginning at the Industrial Revolution, in which humans have substantially altered Earth's ecosystems. To place the words 'good' and 'Anthropocene' together, even with the former in quotation marks, is heresy to some. In *Scientific American* on 19 June, the Australian ethicist Clive Hamilton delivered a broadside against those who argue that human ingenuity, not behaviour change, is the best response to rising carbon emissions. "Such unbounded optimism is dangerous, wishful thinking," he wrote, because it plays into the hands of those who would protect the status quo, whatever the environmental consequences.

Perhaps there is more to this than division along the optimism–pessimism axis; whether someone is the sort of person who sees an atmosphere half-full or half-empty of carbon dioxide. As is made clear in a useful report published this week by the UCL Policy Commission on Communicating Climate Science in London, public discussion of climate change has deeper psychological roots. It is "as much about what sort of world we wish to live in, and hence about ethics and values, as it is about material risks to human wellbeing".

The report is, for want of a better word, ecopragmatic. It acknowledges an awkward truth: if climate scientists are failing to get their message across to the wider public, then it could be the message that is the problem, not the public. The authors of the report acknowledge that some of their conclusions are controversial — scientists must learn to tell stories rather than report cold facts, for instance — but they deserve discussion. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

© 2014 Macmillan Publishers Limited. All rights reserved



Rural students are being left behind in China

As the education gap between city and countryside widens, young people face an invisible barrier to scientific research, says **Qiang Wang**.

People in China have used education to boost their social standing and economic status for centuries. The belief that the system is fair and meritocratic is enshrined in the long-standing motto: "Knowledge can change fate."

Yet for many of the 9.4 million Chinese teenagers who completed their high-school graduation exams earlier this month, the situation is not so simple. A damaging divide has opened between the academic performance of young people from urban and rural areas. This is unfair, restricts opportunity and could stop promising talent from heading into science.

About two-thirds of the students who took the school leaving exam were from the countryside in 2010. But in the same year, less than one in five of new students at Tsinghua University in Beijing — a member of China's elite C9 league of universities — were from rural areas. And the proportion of rural students at another C9 institution, Peking University in Beijing, has decreased from roughly 30% to 10% in the past decade.

Instead, most students from the countryside are shunted into lower-quality colleges. For example, 82% of students in medium-ranking vocational colleges were from rural areas in 2012.

If the highest-quality research universities are so dominated by urbanites, there will be a knock-on effect for science. Just like gender bias in science and engineering careers, rural students will be blocked from postgraduate study and scientific research. And this invisible barrier is amplified by the 'Matthew effect': the sociological phenomenon whereby the rich get richer and the poor get poorer.

Indeed, some professors at Chinese universities already talk of postgraduates from rural areas being less creative and worse at foreign languages than urban rivals with identical test scores. As a direct result, opportunities for rural students to engage in science as a career are reducing, along with the national output of Chinese researchers from the countryside.

If the trend we have seen over the past decade continues, education will lose its role as a social leveller and a safety valve that enables talented but poor youths to escape poverty in China. This would be a disaster for a country that can boast perhaps human society's first attempt to promote upward social mobility and attain social equality through education.

Since AD 607, China has used official examination results to select people to work in the state bureaucracy with no regard for family background or wealth. The *keju* exam provided an open and fair channel for the Chinese populace to fundamentally improve their livelihood and that of their family. This has long influenced Chinese attitudes to education and can partly explain why the Chinese so emphasize the importance of school work.

The *keju* no longer exists, but its spirit survives in the annual *gaokao* exam, established in 1952. The exam is strictly invigilated, and its scoring process is blind. Chinese universities traditionally select students for admission using only *gaokao* results.

A Chinese proverb vividly describes the process of changing destiny through hard learning as "carp leaping over the dragon's gate". Preparing for the 'jump' of the *gaokao* defines the early lives of many young people in China, especially in rural areas.

In the past decade, fewer rural students have made that jump. The main culprit is the unequal distribution of educational resources in rural and urban areas, which starts in preschool education. Children in cities begin to learn foreign languages (one of the three mandated subjects of the *gaokao*) in kindergarten, whereas rural primary schools struggle even to find English teachers.

MOST STUDENTS FROM THE **COUNTRYSIDE** ARE SHUNTED INTO **LOWER-QUALITY** COLLEGES.

Public investment in education is much lower in rural areas than in cities. For example, in 2011, urban primary schools received 700 renminbi (US$112) more public funding per student than did rural primary schools, and the gap was 900 renminbi per student for junior high schools. Even worse, leading Chinese universities have adopted policies to admit more students from their own cities in the name of supporting local development, although most of their funding is from central government and thus their duty is to serve all citizens nationwide.

I propose action in three areas. First, China must accept that the different approaches taken in rural and urban areas are a problem and must be unified. For example, official procedures for evaluating schools and selecting staff should be standardized. Second, China needs to redistribute educational resources. The government should allocate more public money to rural education and allow students from the countryside to take their college entrance examinations in urban areas. Finally, charities and other non-governmental groups could help by encouraging graduates from prestigious Chinese and Western universities to take up teaching positions in rural areas. Most such areas, especially in western China, have a severe shortage of teachers, let alone native English speakers.

So far, the Chinese people seem willing to accept wide inequality in income as the price of economic growth. Inequality in education is less palatable and could trigger social unrest. It should be tackled. ■

➲ **NATURE.COM**
Discuss this article online at:
go.nature.com/pq8xlb

Qiang Wang *is Conjoint Professor at the Xinjiang Institute of Ecology and Geography of the Chinese Academy of Sciences in Urumqi.*
e-mail: qiangwang7@gmail.com

© 2014 Macmillan Publishers Limited. All rights reserved

RESEARCH HIGHLIGHTS

Selections from the scientific literature

CARDIOVASCULAR BIOLOGY

Mutations lower heart-disease risk

Two groups have found rare gene mutations linked to reduced blood fat levels and a lower risk of heart disease.

The *APOC3* gene codes for a protein that increases blood triglyceride levels — a known risk factor for heart disease. So, Anne Tybjaerg-Hansen at the Copenhagen University Hospital and her colleagues sequenced the protein-coding regions of the gene in more than 10,000 people and found three rare mutations associated with low blood triglyceride levels. They analysed these mutations in an additional 65,000 people and found that individuals with any of the mutations had a 36–41% lower risk of heart disease and stroke than people without any of the mutations.

Sekar Kathiresan at the Broad Institute in Cambridge, Massachusetts, and his team sequenced the protein-coding regions of nearly 20,000 genes from more than 3,500 participants and came up with four mutations in the same gene with similar effects.

The findings suggest that lowering blood triglyceride levels could protect against heart disease.

N. Engl. J. Med. **http://doi.org/s98; http://doi.org/s97 (2014)**



RICHARD ELLIS/ALAMY

ANIMAL BEHAVIOUR

Magnetic compass guides butterflies

On overcast days, monarch butterflies use a magnetic compass to find their way south, making them one of only a few migratory insects known to sense Earth's magnetic field.

The eastern North American monarch butterflies (*Danaus plexippus*; pictured) use the Sun to guide them from southern Canada and the United States towards Mexico, but they still manage to fly in the correct direction on cloudy days. Steven Reppert at the University of Massachusetts Medical School in Worcester and his team studied butterflies in a flight simulator inside an artificial magnetic field. The authors found that the butterflies changed their flight orientation when the researchers shifted the magnetic field, but only when the insects were exposed to certain wavelengths of light.

The insects' antennae could contain magnetosensors, the authors say, adding that human-induced electromagnetic noise might disrupt the butterflies' migration.

Nature Commun. **http://dx.doi.org/10.1038/ncomms5164 (2014)**

NEUROSCIENCE

Brain circuit spurs social behaviour

Neuroscientists have pinpointed a specific set of neurons that connects two brain areas and regulates social behaviour in mice.

Karl Deisseroth and his colleagues at Stanford University in California used a variety of technologies to identify the exact circuitry involved in the behaviours adopted by mice encountering an unfamiliar mouse or object. The team found that activity of cells in the ventral tegmental area (VTA) region of the brain increased during exploration of both the mouse and the object. Stimulating the firing of the VTA cells that project into the nucleus accumbens increased the animals' exploration of their new companion — but not the object. The mice also seem to use this circuitry for natural social exploration.

Studying the brain circuits, and not just the molecules, involved in social behaviour could lead to new insight into some neuropsychiatric disorders, the authors say.

Cell 157, **1535–1551 (2014)**

ATMOSPHERIC SCIENCE

Arctic heat lessens cold snaps

The middle latitudes of the Northern Hemisphere have experienced fewer cold snaps during autumn and winter, thanks to Arctic warming.

Earlier studies have linked unusually frigid winters in North America and Europe over the past decade to changes in atmospheric circulation caused by rising Arctic temperatures. But when James Screen of the University of Exeter, UK, analysed global temperature data from 1979 to 2013, he found that the frequency and severity of autumn and winter cold extremes decreased during this period. In these seasons, the coldest days in the Northern Hemisphere's high and middle latitudes have warmed faster than the warmest days, reducing overall temperature

© 2014 Macmillan Publishers Limited. All rights reserved

variation. Climate models suggest that this trend will continue in the future.

The rapid warming of northerly winds, which carry cold Arctic air south, could be dampening winter extremes, Screen says.

Nature Clim. Change **http://doi.org/s88 (2014)**

ORGANIC CHEMISTRY

Microbes pitch in with synthesis

The metabolism of living organisms could be harnessed to help construct small molecules, according to a team from Harvard University in Cambridge, Massachusetts.

Chemists routinely use microbial enzymes as catalysts. But Emily Balskus and her colleagues instead used engineered *Escherichia coli* as a source of reagent. They combined a palladium-based catalyst with the hydrogen generated by the *E. coli* to add hydrogen to a variety of alkenes (which contain double-bonded carbon atoms).

Although the reaction was not as efficient as conventional methods, the approach could produce molecules that cannot be made using biochemistry alone, the researchers argue.

Angew. Chem. **http://doi.org/f2sbt2 (2014)**

LEE CHRISTENSEN/MOMENT/GETTY

CLIMATE SCIENCE

Warming could boost air pollution

Climate change could lead to increased air pollution in the most populated areas, thanks to greater stagnation of air masses allowing pollutants such as ground-level ozone to build up.

A team led by Daniel Horton of Stanford University in California used climate models to study how global warming can produce stagnant air by altering atmospheric circulation and precipitation patterns. The models predict that air stagnation will increase over the rest of this century across most of the tropics and subtropics. For example, large swathes of India (**pictured**) and Mexico could experience more 'stagnation days', and longer lasting stagnation events.

The areas at highest risk are home to roughly 55% of the world's population.

Nature Clim. Change **http://doi.org/tch (2014)**

ERIC REED/ZUMA PRESS/CORBIS



ASTRONOMY

Speedy stars revealed nearby

Astronomers have spotted 28 stars that are hurtling through space fast enough to escape the Milky Way's gravitational pull — the biggest set of such stars, and the nearest to Earth, so far identified.

A team led by Jing Zhong of the Shanghai Astronomical Observatory, China, analysed data on star speeds gathered by the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) in Hebei province. The scientists found 28 stars moving faster than 300 kilometres per second with respect to the galaxy. Some were within 1,000 parsecs of Earth.

High-velocity stars are thought to gain their speed through gravitational interactions with nearby stars or black holes — finding so many of these stars could reveal which mechanisms are most common.

Astrophys. J. Lett. 789, **L2 (2014)**

ANTHROPOLOGY

To and from the Horn of Africa

After emerging from Africa 100,000 years ago, humans migrated back to the continent earlier than previously thought.

A team led by Ryan Raaum at the City University of New York in the Bronx compared nuclear genome data from several populations living in the Horn of Africa with those of humans from across the Middle East and elsewhere. Groups such as Ethiopians and Somalians have non-African ancestry that is distinct from that of present-day people of the Middle East, indicating that interbreeding between African and Middle Eastern people occurred more than just a few thousand years ago, as other studies have suggested.

The authors conclude that humans were moving back to the Horn of Africa as early as 23,000 years ago, before the advent of agriculture.

PLoS Genetics 10, **e1004393 (2014)**

SOCIAL SELECTION

Popular articles on social media

Secret publishing deals exposed

Judging by the social-media reaction, researchers worldwide are paying attention to a report by US economists detailing the confidential deals that universities make with big publishers for electronic access to multiple journals. The prices, which can exceed US$1 million a year, vary hugely between institutions, and do not always represent a good buy. For instance, libraries that buy bundled deals from Elsevier (which sells more than 2,200 journals) pay about three times more per citation than they do for non-profit items such as the *Journal of the American Chemical Society*. (Nature Publishing Group was not included in the study.) "Large commercial publishers offer v. poor value," tweeted biologist Ross Mounce of the University of Bath, UK.

Proc. Natl Acad. Sci. USA **http://doi.org/s78 (2014)**

Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

NATURE.COM
For more on popular papers:
go.nature.com/u5kxm3

NEUROSCIENCE

Tanning might be addictive

Repeated visits to the beach or tanning salon could be signs of an addiction, according to a study in mice.

David Fisher at the Massachusetts General Hospital in Boston and



his team found that mice chronically exposed to ultraviolet light produced an opioid called β-endorphin, which numbs pain and is associated with addiction. The mice also showed signs of withdrawal, including shaky paws and chattering teeth, after treatment with an opioid-blocking drug. The effects of β-endorphin, however, were not as pronounced as those seen in earlier studies of mice given morphine.

If the results hold true in humans, they could explain why many tanners (**pictured**) continue to seek out the sun, despite the known risks.

Cell 157, **1527–1534 (2014)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

© 2014 Macmillan Publishers Limited. All rights reserved

SEVEN DAYS *The news in brief*

POLICY

Marine reserves

US President Barack Obama said on 17 June that he wants to expand the Pacific Remote Islands Marine National Monument, near Hawaii, from 225,000 square kilometres to around 2 million square kilometres. Meanwhile, another large marine reserve, the Phoenix Islands Protected Area in Kiribati, will close to commercial fishing by the end of the year, Kiribati's President Anote Tong said on 16 June. The area is an important spawning ground for tuna.

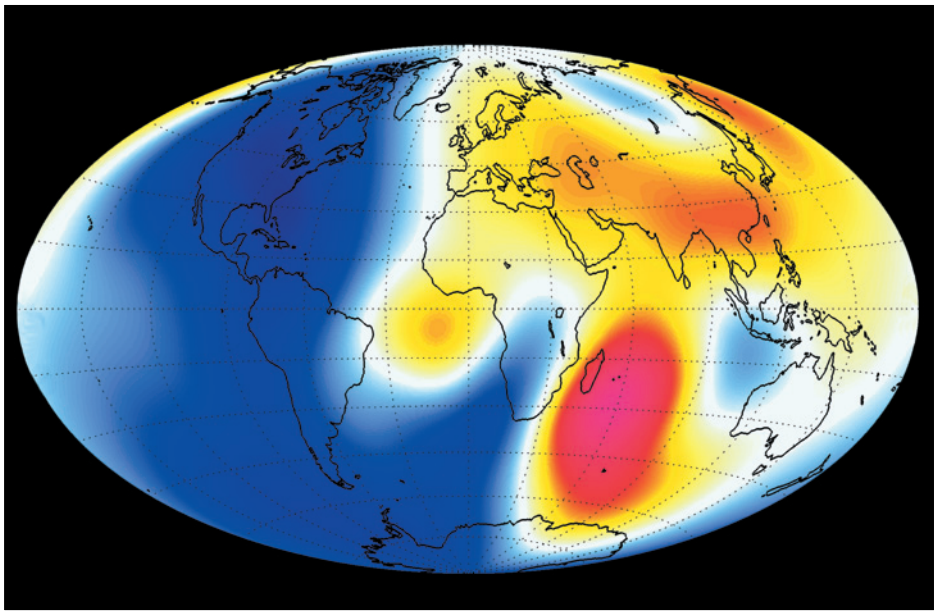
PEOPLE

Research fraud

A scientist who faked HIV vaccine data was charged by US prosecutors on 19 June with making false statements. Dong-Pyou Han, a former researcher at Iowa State University in Ames, had already been found guilty of misconduct by the US Department of Health and Human Services' Office of Research Integrity and banned from working with US government agencies for three years, beginning in November 2013. Han was found to have added antibodies to blood from rabbits that had been given an experimental HIV vaccine, making it seem as if the vaccine had generated an immune response against the virus.

Chemist fined

Patrick Harran, a chemist at the University of California, Los Angeles, has been ordered to pay US$10,000 to the burns unit where research assistant Sheharbano Sangji was treated but died in 2009 after a chemical fire while working in Harran's lab. The fine is part of a deferred-prosecution deal reached on 20 June, which will probably spare Harran a public trial and jail time. Harran was charged in 2011 with "willful violation of an occupational safety and health standard causing the death of an employee". See go.nature.com/vpdwvz for more.



DTU SPACE/ESA

Satellites show magnetic field in decline

The first images from Europe's three-satellite Swarm constellation show that Earth's magnetic field is weakening on average by about 5% every century. The magnetic field shields Earth from charged particles emitted by the Sun. The blue areas in the image, released by the European Space Agency on 19 June, show where the field is weakening, most dramatically, in the Western Hemisphere. The yellow and red areas indicate where the field is strengthening — notably, over the Indian Ocean. The measurements, collected from January to June 2014, also confirm the movement of the north magnetic pole from Canada across the Arctic towards Siberia. Scientists will continue to analyse the data to understand why the magnetic field is weakening.

RESEARCH

Vanishing waves

Scientists who claimed to have seen gravitational waves in the Big Bang's afterglow now admit that dust may be getting in the way (P. A. R. Ade *et al. Phys. Rev. Lett.* **112,** 241101; 2014). In March, a team operating the BICEP2 telescope at the South Pole said that it had seen evidence of gravitational ripples from a period of rapid expansion just after the Big Bang (see *Nature* **507,** 281–283; 2014). But critics said that the telescope might have been confused by dust floating inside the Milky Way. See go.nature.com/f6fnbo for more.

HIV prevalence

Epidemics of HIV infection in people who inject drugs are flourishing in one-third of the 23 Middle Eastern and North African countries, finds the first large study on this region. The data, published on 17 June, show that, in some places, HIV prevalence among injecting drug users is among the highest in the world (G. R. Mumtaz *et al. PLoS Med.* http://doi.org/tcf; 2014). For example, in Tripoli, Libya, more than 87% of injecting drug users are infected. There are around 2 injecting drug users per 1,000 adults in the region — a comparable figure to other parts of the world. High-risk behaviour, such as sharing needles, and low rates of condom use could be fuelling the surge, the study says.

FUNDING

Graphene boost

Europe's Graphene Flagship collaboration doubled in size on 23 June, when the European Commission announced that 66 new partners would join

© 2014 Macmillan Publishers Limited. All rights reserved

the €1-billion (US$1.4-billion) initiative. Funding awards for 2014–15 were also announced, with €9 million to be shared by 21 winning proposals. Public and private researchers will work together in the decade-long initiative to commercialize graphene, a one-atom-thick carbon material.

FACILITIES

Satellite record

The Russian rocket Dnepr carried 33 satellites into space on 19 June, a record for a single launch. It hauled payloads from 17 nations, including remote-sensing spacecraft for Spain and Kazakhstan; a joint US–Saudi Arabian mission to test technology for gravitational-wave detectors; and two Japanese probes that will monitor the environment around nuclear power plants. The company Planet Labs in San Francisco, California, sent up 11 craft for its planned network of 100 Earth-imaging satellites.

EVENTS

Ebola outbreak

West Africa is in the grip of the largest-ever outbreak of disease caused by the Ebola virus, the World Health Organization said on 23 June. Thirty-eight new cases and 16 deaths have been reported from Guinea (**pictured**) and Sierra Leone

SEYLLOU/AFP/GETTY



since 14 June, and 19 new cases and 13 deaths from Liberia since 11 June. This brings the official total to 599 confirmed or suspected infections and 338 fatalities since the outbreak began in March. This is the first known Ebola outbreak in West Africa; the virus was first reported in 1976 in the Democratic Republic of the Congo (then Zaire).

Anthrax exposure

Dozens of scientists working at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, may have been exposed to live *Bacillus anthracis*, the bacterium that causes anthrax, the agency said on 19 June. Scientists did not follow proper procedures to kill the bacteria when preparing them for transportation to other CDC labs. Believing the bacteria harmless, the staff did not wear protective clothing. The labs also may have inadvertently allowed the bacteria to become airborne, directly exposing about seven employees. Both the CDC and the Federal Bureau of Investigation are looking into the incident, and at least 70 employees are being treated with antibiotics. See page 443 for more.

BUSINESS

Patent ruling

The US Supreme Court moved to limit software patents in a ruling issued on 19 June. The case, *Alice Corporation* v. *CLS Bank International*, was a lawsuit over patents on programs to reduce financial risk in certain transactions. The Supreme Court said that the patents were invalid because they represented abstract ideas, which are not patentable. The decision could limit broad software patents that are fodder for patent trolls, and could have implications for other kinds of patents, including those covering methods for medical diagnosis (see *Nature* **507,** 410–411; 2014).

Personal genetics

The genetic-testing company 23andMe in Mountain View, California, is trying to get regulatory approval for a test it intends to sell direct to consumers. On 20 June, the firm said that the US Food and Drug Administration (FDA) had accepted its application for a report on Bloom syndrome, an inherited condition that causes skin problems, short stature and elevated risk of cancer. The FDA had told 23andMe to stop marketing its tests because the company had not provided information showing that its DNA service is safe and effective.

COMING UP

30 JUNE–4 JULY
The European Week of Astronomy and Space Science is taking place in Geneva, Switzerland.
go.nature.com/y9tm64

2–9 JULY
Physicists meet at the 37th International Conference on High Energy Physics in Valencia, Spain.
go.nature.com/226uku

AWARDS

Kyoto prizewinners

This year's ¥50-million (US$500,000) Kyoto prizes in science have been won by biomedical engineer Robert Langer and mathematician and theoretical physicist Ed Witten. Langer was honoured for pioneering tissue engineering and Witten for advancing superstring theory. The Inamori Foundation in Tokyo announced the awards on 20 June.

Maths prize

Five people were awarded the inaugural Breakthrough Prize in Mathematics on 23 June. The awards — worth US$3 million each — were given for work largely at the intersection of pure maths and theoretical physics. They are sponsored by billionaire philanthropists, including Internet entrepreneur Yuri Milner and Facebook co-founder Mark Zuckerberg. See go.nature.com/yijkrc for more.

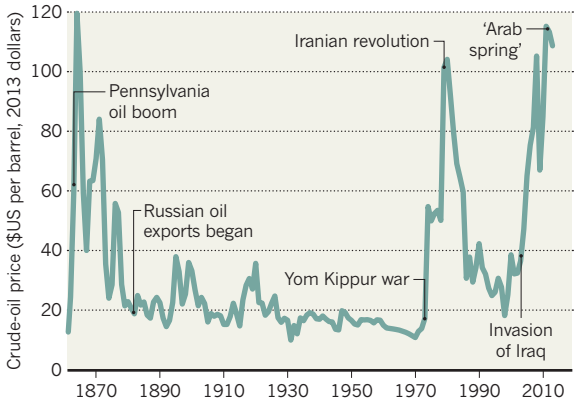
NATURE.COM
For daily news updates see:
www.nature.com/news

TREND WATCH

Oil remains the world's dominant energy source, according to the *BP Statistical Review of World Energy*, released last week. Its share has declined steadily, however: from close to 50% of energy consumption by raw (or 'primary') fuel source in the 1970s, to 32.9% in 2013. And its price has remained above US$100 a barrel for three years, owing to disruptions to supply in the Middle East and Africa. Non-fossil sources crept up to 13.3% of primary energy consumption last year, equal to 1999 levels.

THE HIGH PRICE OF OIL

Despite huge US production increases, average oil prices are at their highest — in real terms — since the Pennsylvania oil boom of the 1860s.



SOURCE: BP

© 2014 Macmillan Publishers Limited. All rights reserved

NEWS IN FOCUS

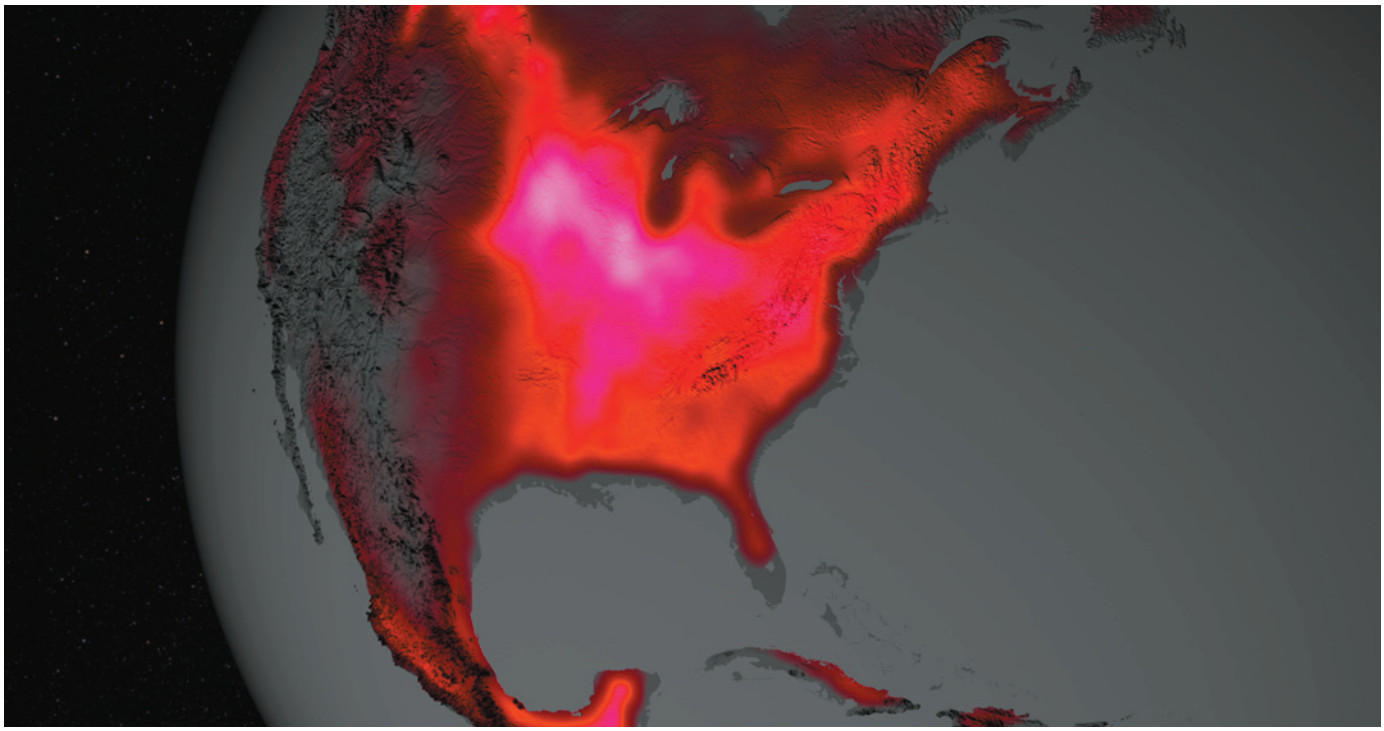
POLICY Pan-African science strategy criticized over lack of commitments **p.452**

DRUG DEVELOPMENT Reanalysis leads to comeback for failed cancer medicine **p.454**

ECOLOGY Project to protect biodiversity kicks off with pollinator study **p.455**

NUTRITION Vitamin research plagued by poor study design **p.462**

NASA GODDARD SPACE FLIGHT CENTER



A fluorescent glow emitted by plants during photosynthesis is picked up by satellite imaging, emphasizing high crop productivity in the US Midwest.

CLIMATE SCIENCE

NASA carbon-monitoring orbiter readies for launch

Satellite will map sources and sinks of greenhouse gas in unprecedented detail.

BY LAUREN MORELLO

If everything goes to plan, on 1 July NASA will attempt to launch its first carbon-monitoring satellite — for the second time.

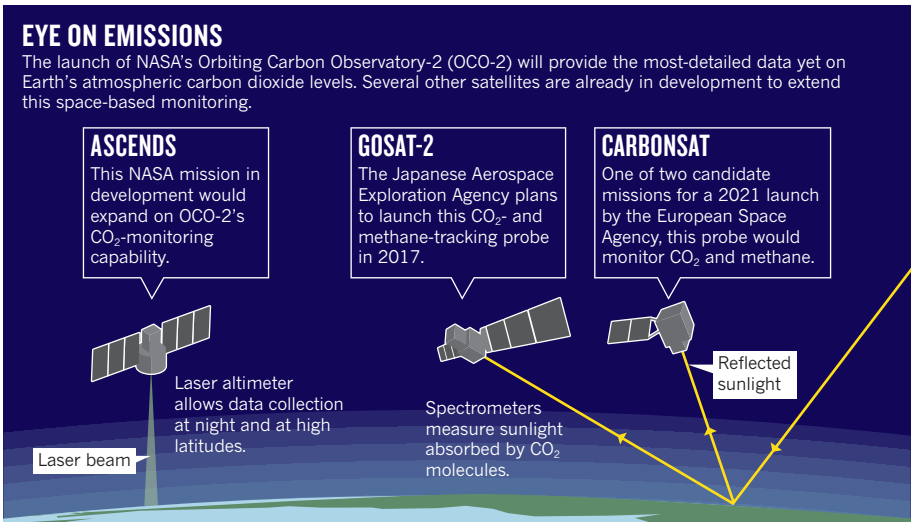
The US$465-million Orbiting Carbon Observatory-2 (OCO-2) is designed to collect precise, high-resolution measurements of the carbon dioxide in Earth's atmosphere. It is a near-copy of a NASA probe that crashed in 2009 before it reached orbit, devastating the scientists who had hoped to use its data to map sources and sinks of the greenhouse gas — both man-made and natural — in unprecedented detail.

Once OCO-2 reaches orbit, its spectrometer will sample a column of air extending from the upper atmosphere down to Earth's surface, taking hundreds of thousands of measurements each day and covering the entire globe every 16 days. Although it will not be the first probe to track CO_2 from space, the satellite will do so at a higher resolution than its predecessors. Each OCO-2 reading will sample a column of the atmosphere above roughly 3 square kilometres of Earth's surface — much smaller than the 85-square-kilometre footprint of a similar Japanese probe, the Greenhouse Gases Observing Satellite (GOSAT), which launched in 2009 and is still operating.

This unprecedented resolution allows OCO-2 to measure CO_2 more accurately in cloudy areas such as the tropics, says Annmarie Eldering, OCO-2's deputy project scientist at NASA's Jet Propulsion Laboratory in Pasadena, California. Others hope that it will also allow OCO-2 to go beyond large natural sources and sinks, such as oceans and forests, to detect CO_2 emanating from sprawling urban areas or even large power plants.

"We should clearly be able to see large metropolitan 'super-cities' like Los Angeles," says Paul Wennberg, an atmospheric ▶

© 2014 Macmillan Publishers Limited. All rights reserved



scientist at the California Institute of Technology in Pasadena who helped to develop the network of ground-based monitoring stations that will be used to test the quality of OCO-2's data. He hopes that the satellite will be able to confirm ground-based measurements that show a small increase in CO_2 levels above Pasadena each day as emissions from the greater Los Angeles area pool in the low atmosphere.

Such data, combined with ground monitoring and inventories of fossil-fuel consumption, could be a powerful tool for pinpointing greenhouse-gas sources, says Kevin Gurney, a carbon biogeochemist at Arizona State University in Tempe. "When we talk to policy-makers they say, 'One number for the whole city is lovely, but we need to know what is emitting'."

OCO-2 will also closely monitor the carbon uptake of plants by measuring the weak fluorescence that is produced during photosynthesis as plants' chlorophyll pigments absorb light to capture energy and subsequently re-emit photons at longer wavelengths. Researchers have already tested the method using observations from GOSAT and the Global Ozone Monitoring Experiment 2 (GOME-2) instrument aboard the European meteorological satellite MetOp-A — with surprising results. An analysis of GOME-2 data published in April suggests that carbon-cycle models underestimate peak photosynthetic output by as much as 50–75% in parts of India, China and the African Sahel, and by 40–60% in the 'corn belt' of the US Midwest, which accounts for more than 40% of the world's maize (corn) production (L. Guanter *et al. Proc. Natl Acad. Sci. USA* **111,** E1327–E1333; 2014).

The results "jumped out at us right away", says Joanna Joiner, an atmospheric physicist at the NASA Goddard Space Flight Center in Greenbelt, Maryland, and co-author of the study. Joiner, who helped to develop the space-based method for monitoring plant fluorescence, is eager to take measurements using OCO-2, the small footprint of which should allow researchers to detect fluorescence signals from plants in more fragmented landscapes such as Europe's. Analyses by GOSAT and GOME-2 have been limited to vast expanses such as North America's boreal forests, the Amazon rainforest and the US corn belt.

OCO-2 is designed to last just two years, although NASA has loaded the probe with enough fuel to keep it running for 10 to 12 years. The space agency is developing another CO_2-observing satellite, the Active Sensing of CO_2 Emissions over Nights, Days and Seasons (ASCENDS) mission, to follow OCO-2. The multifrequency laser on ASCENDS would allow researchers to measure atmospheric CO_2 at night and at high latitudes in all seasons, something that the spectrometers on GOSAT and OCO-2 cannot manage, because they require sunlight to detect the characteristic absorption signal of CO_2 (see 'Eye on emissions'). Both Japan and Europe are also developing next-generation CO_2-monitoring satellites.

NASA hopes to install a duplicate of the OCO-2 instrument (originally built as a flight spare) on the International Space Station in 2017, at a cost of roughly $120 million. The station's low-Earth orbit would allow the spectrometer — dubbed OCO-3 — to gather data at different times of day from OCO-2, which follows a Sun-synchronous path and thus passes over areas in their early afternoons. Wennberg says that combining such measurements with OCO-2 data could help researchers to detect hourly variations in CO_2 production caused by plant photosynthesis — or even by rush-hour traffic in some large urban areas.

Plans for OCO-3 were dealt a blow in March, when US President Barack Obama proposed a NASA budget that did not include funding for the carbon-sensing project. But the agency says that it could revive OCO-3 if money becomes available. "We hope to get back to OCO-3," says Betsy Edwards, programme executive for OCO-2 at NASA in Washington DC. "We are putting it into a position where we could do that at any time." ■

POLICY

Africa science plan attacked

Proposed innovation strategy is low on detail and commitments from governments.

BY LINDA NORDLING

Scientists have raised concerns about a ten-year pan-African science and innovation strategy that heads of state are expected to adopt this week.

The Science, Technology and Innovation Strategy for Africa (STISA), proposed by the African Union (AU) continental grouping, aims to prioritize the use of research to drive economic and social development across the continent. It will commit signatory countries to six goals, including tackling hunger, disease and unemployment, and will set up structures that give research and development organizations leverage to pursue them.

But critics fear that the strategy's top-heavy administrative structure and lack of firm pledges may render it ineffective. They also think that its aims may be beyond the continent's limited resources, especially given that it contains few financial commitments.

"The majority of the document talks about governing structures and their responsibilities, and very little about implementation," says Paul van Gardingen, who studies international development at the University of Edinburgh, UK. "It's like trying to reach Mars whilst talking about who is responsible for designing the hot-air balloon to take you there."

© 2014 Macmillan Publishers Limited. All rights reserved

Africa's research-publication output is small, but growing faster than the world average. The share of global scientific papers with at least one African author went from 1.2% a year in 1996 to 2.3% in 2012, with the number of such papers quadrupling from about 12,500 a year to 52,000.

STISA, which is likely to be signed off during this week's AU summit in Malabo, Equatorial Guinea, aims to turn this intellectual capacity into a "tool and enabler for achieving continental development goals". But the document, drafted by the AU and a group of African science policy-makers, does not give much detail on how this vision will be achieved.

It mentions the need to invest in new laboratories and other infrastructure, expand graduate and technical training programmes and create an "enabling environment" for technology transfer. But it does not commit governments to concrete spending or training targets. Nor does it provide a budget, indicating only that funds need to be sought from both inside and outside Africa. It reiterates a near-decade-old ambition to raise AU member states' spending on research and development to 1% of gross domestic product — a target that few have achieved (see 'Continental divide').

Success of the strategy will depend on the quality of research projects in individual countries. To help scientists to gain domestic support for research programmes, STISA plans to set up a research and innovation council that will bring together academies and funders to coordinate national activities. It will also take control of a European Union-funded competitive granting scheme that has spent almost €14 million (US$19 million) on research projects in water and sanitation, agriculture and energy.

Critics ask whether the new council is necessary. "The current structures look very top-heavy and could be far too expensive for the limited investments likely to be realized in the early years," says van Gardingen.

John Mugabe, a technology-studies specialist at the University of Pretoria in South Africa, is also sceptical. "I would prefer the use of existing institutions such as the African Academy of Sciences and national academies to mobilize technical expertise for programme design and implementation," he says. But Abdul-Hakim Elwaer, director of administration and human-resources management at the African Union Commission in Addis Ababa, Ethiopia, notes that not all African countries are involved in the African Academy of Sciences, whereas all AU member states will be represented by the council.

Despite their concerns, critics agree that STISA is an improvement on its predecessor, Africa's Science and Technology Consolidated Plan of Action (CPA). Launched in 2005 and adopted by the AU in 2006, it established the competitive research grants scheme and several networks of research excellence. But a review in 2013 found that the CPA had failed to link with other pan-African policies, such as continent-wide agriculture and environmental-protection projects with research programmes of their own. It also fell short of raising the funds needed for its full implementation.

Mugabe, who helped to spearhead the CPA, expects funding to be a problem for STISA, too. "It fails to address the big issue, which is the financing of science, technology and innovation in Africa by African governments," he says. The absence of commitments mean that heads of state will probably adopt it without much debate, he adds, and governments will be able to ignore it easily thereafter.

Daan du Toit, head of international cooperation and resources at South Africa's Department of Science and Technology, is more positive. He says that the CPA generated political support for science that paid off in the continent's successful bid to co-host the Square Kilometre Array, a US$2-billion radio-telescope project being built in South Africa and Australia.

The latest strategy will provide a similar political basis on which to build future partnerships, he says. "STISA will be a very strong mandate. You can squabble about the detail, but the overall vision is most important." ■

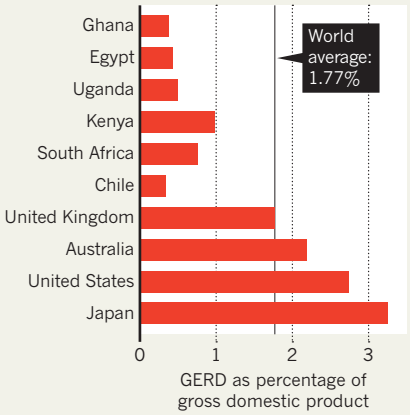
SVEN TORFINN/PANOS PICTURES



Crop research is one of several sectors set to benefit from a proposed African science strategy.

CONTINENTAL DIVIDE

A sample of leading African science spenders shows that they lagged behind many major economies in gross domestic expenditure on research and development (GERD) in 2010.

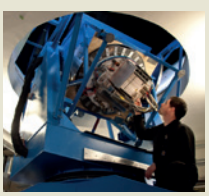


SOURCE: AFRICAN INNOVATION OUTLOOK II/OECD

STEFFEN RICHTER/HARVARD UNIV.

MORE ONLINE

TOP STORY



Gravitational-wave paper is published — with caveats go.nature.com/f6fnbo

MORE NEWS

- Egg patterning allows birds to identify invading cuckoos go.nature.com/28mjfc
- Higgs seen decaying into matter particles go.nature.com/flonak
- Europe's science chief seeks expert voices go.nature.com/c7asyf

NATURE PODCAST



Cheaper solar-cell manufacturing, love in the lab, and how to detect nuclear weapons nature.com/nature/podcast

© 2014 Macmillan Publishers Limited. All rights reserved

DRUG DEVELOPMENT

Resurrected cancer drug faces regulators

Despite a chequered history, olaparib is finally before the US Food and Drug Administration.

BY HEIDI LEDFORD

A cancer drug once lauded as a shining example of translational research but cast aside after a disappointing clinical trial may be rising from the ashes.

Olaparib was one of the first drugs to target enzymes that repair broken strands of DNA. This week it is facing scrutiny by a US Food and Drug Administration (FDA) advisory panel, in a meeting that could shape the agency's decision over whether to approve the drug later this year for use in a subset of ovarian cancers.

The maker of the drug, AstraZeneca in London, is not the only party eagerly awaiting the verdict. Academic researchers who think olaparib still has merit and a handful of companies that are developing similar drugs are also on high alert. Although the disappointing results led AstraZeneca to halt clinical tests of olaparib in 2012, a reanalysis[1] of data from the largest completed trial[2] has reignited interest in the class of drug — known as poly(ADP-ribose) polymerase, or PARP, inhibitors.

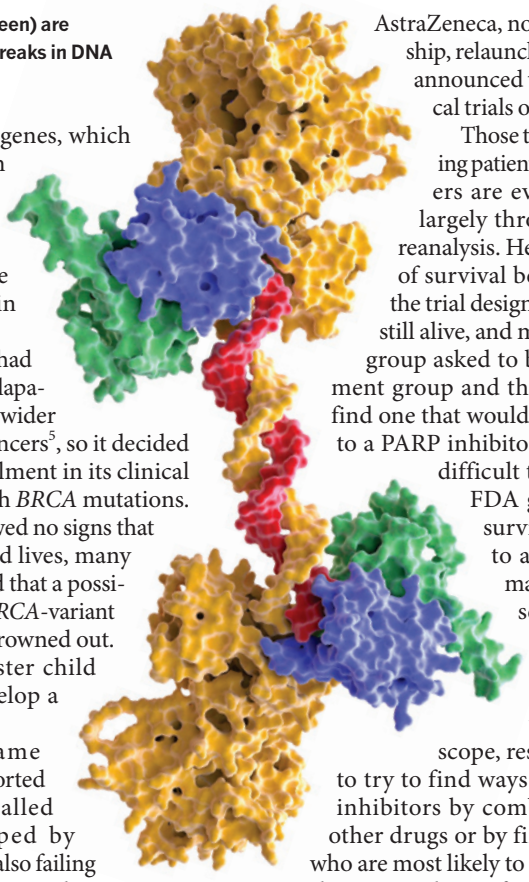
"Two and a half years ago, this drug was dead," says Michael Birrer, an oncologist at Massachusetts General Hospital in Boston. "Now it's: 'Here a PARP, there a PARP, everywhere a PARP-PARP'."

GOLDEN OPPORTUNITY

Olaparib is the product of almost half a century of research showing that PARPs help to mend DNA damage. Left unmended, breaks in both strands of DNA's double helix can trigger cell death. Inhibiting the enzymes' action is unlikely to kill healthy cells, because they have multiple pathways to fix broken DNA. But cancer cells sometimes have mutations that knock out other types of repair, making them particularly sensitive to PARP inhibition. So a drug with this mechanism would target cancer cells and bypass healthy ones, avoiding some of the toxic side effects of conventional chemotherapy.

"It's become a poster child in how not to develop a drug."

Studies done in mice[3] and cells[4] suggested that PARP inhibitors would be most effective in patients who carry specific variants of the *BRCA1* or *BRCA2* genes, which are associated with some aggressive forms of breast and ovarian cancer and encode proteins involved in DNA repair.



ALFRED PASIEKA/SPL

PARPs (purple and green) are enzymes that repair breaks in DNA (red and yellow).

But AstraZeneca had seen evidence that olaparib might combat a wider range of ovarian cancers[5], so it decided not to restrict enrolment in its clinical trial to patients with *BRCA* mutations. When the trial showed no signs that olaparib lengthened lives, many researchers believed that a possible benefit against *BRCA*-variant cancers had been drowned out. "It's become a poster child in how not to develop a drug," says Birrer.

Around the same time, another purported PARP inhibitor called iniparib, developed by Sanofi in Paris, was also failing in clinical trials. Researchers would later show that iniparib was not a true PARP inhibitor, but by then, interest in the drug class had been sapped (see *Nature* **483,** 519; 2012). "It dirtied the waters," says Scott Kaufmann, an ovarian-cancer researcher at the Mayo Clinic in Rochester, Minnesota.

AstraZeneca halted its PARP programme, and other major pharmaceutical companies sold off theirs. The promise of a less-toxic treatment for ovarian cancer seemed to have vanished. "We all just groaned," says Birrer. "And the patients were screaming."

That changed when oncologist Jonathan Ledermann of the University College London Cancer Institute reanalysed the trial data, this time focusing on patients with cancer-linked *BRCA1* or *BRCA2* mutations. The results, announced last year and published in May[1], showed that although olaparib did not lengthen survival in patients with the mutations, it did slow cancer growth.

AstraZeneca, now with new leadership, relaunched its research and announced two late-stage clinical trials of the drug.

Those trials are still recruiting patients, so the FDA advisers are evaluating olaparib largely through Ledermann's reanalysis. He attributes the lack of survival benefit to factors in the trial design: many patients are still alive, and many in the placebo group asked to be told their treatment group and then left the trial to find one that would allow them access to a PARP inhibitor, making the data difficult to analyse. But the FDA generally requires survival improvements to approve a drug, so many are interested to see how the agency will weigh up those explanations.

As olaparib goes under the microscope, researchers continue to try to find ways to improve PARP inhibitors by combining them with other drugs or by finding the patients who are most likely to benefit from them.

Clovis Oncology of Boulder, Colorado, is now developing a way to identify patients whose cancers show evidence of faulty DNA repair, even if they do not have mutations in *BRCA1* and *BRCA2*. If these patients respond to the inhibitor it is developing, called rucaparib, it could broaden the number of people who will benefit from the drugs.

Despite the dismal prospects two years ago, interest in the field has never been higher, says Guy Poirier, a biochemist at Laval University in Quebec, Canada, who has studied PARPs for 40 years. "I think we're just seeing the beginning." ■

1. Ledermann, J. *et al. Lancet Oncol.* http://dx.doi.org/10.1016/S1470-2045(14)70228-1 (2014).
2. Ledermann, J. *et al. N. Engl. J. Med.* **366,** 1382–1392 (2012).
3. Rottenberg, S. *et al. Proc. Natl Acad. Sci. USA* **105,** 17079–17084 (2008).
4. Farmer, H. *et al. Nature* **434,** 917–921 (2005).
5. Gelmon, K. A. *et al. Lancet Oncol.* **12,** 852–861 (2011).

© 2014 Macmillan Publishers Limited. All rights reserved



JOHN KIMBLER/NATIONAL GEOGRAPHIC

The first in a series of new global biodiversity assessments will focus on pollinators such as bees.

ECOLOGY

'Life on Earth' project gets under way

Assessments by international biodiversity group aim to halt damage to world's ecosystems.

BY NATASHA GILBERT

An international push to protect global biodiversity will take centre stage in Bonn, Germany, next week as researchers gather to discuss its first project — assessing the health of bees and other pollinators.

The Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) in Bonn aims to halt attacks on ecosystems and track the ecological health of the planet. As part of its first scientific meeting, scientists will thrash out the details of the pollinator assessment, which includes analyses of the causes of the decline in bee populations and how it might affect food production.

The review is the first in a series of ambitious biodiversity assessments that IPBES hopes will inform policy decisions. The platform's high-level international leadership makes it more likely that its results will be taken seriously, says executive secretary Anne Larigauderie.

Governments established IPBES in 2012 in response to concerns from the scientific community about the lack of international efforts to tackle threats to biodiversity. It now has 118 member nations. Modelled on the Intergovernmental Panel on Climate Change (IPCC), IPBES brings together scientists and decision-makers to monitor the state of the world's biodiversity and natural resources. There is a sense of urgency to the work; for example, the amount of land lost to soil erosion jumped from 15% in 1991 to 24% in 2008.

One of the platform's key tasks is to assess scientific, social and local knowledge of the diversity of life on Earth, and predict how ecological changes will affect ecosystems and people. Like the IPCC, the platform will not conduct its own research; instead, it will build on existing work, such as the 2005 United Nations Millennium Ecosystem Assessment. IPBES aims to feed its conclusions into policy debates on how to protect ecosystems and pursue sustainable economic growth.

The pollinator study will be followed by one on biodiversity modelling, which will be discussed by a separate group of scientists in the Netherlands in October. It will review existing ways of evaluating ecological diversity, including computer simulations and measures of species abundance used to compile, for example, the Red List of Threatened Species, drawn up by the International Union for Conservation of Nature (IUCN), based in Gland, Switzerland.

The assessment will also explore the need for more complex simulations of ecological processes and their outcomes. These could use economic and social parameters similar to those developed for the IPCC, such as population growth rates or per capita food consumption. That would improve the ability of ecologists to predict future trends in biodiversity, gauge their impact on society and compare different scenarios.

Larigauderie says that IPBES aims to complete its first two assessments by the end of 2015, a goal that she acknowledges is "very ambitious". Future candidates for assessment, to be discussed at an IPBES meeting in January, include land degradation and restoration; regional assessments of regions such as Africa or Europe; and invasive species.

In all its reviews, IPBES will aim to identify gaps in scientific knowledge. It plans to help fill these by holding annual meetings with research organizations and funding agencies to discuss its findings and suggest where future research efforts are needed.

IPBES aims to stand out from other biodiversity assessments — as well as from the IPCC — by incorporating traditional knowledge and social science in its assessments. The former is often overlooked by scientists, says Douglas Nakashima, chief of the indigenous knowledge unit at the Paris-based United Nations Educational, Scientific and Cultural Organization. But there are many examples of such knowledge of local ecosystems turning out to be of great value, he adds.

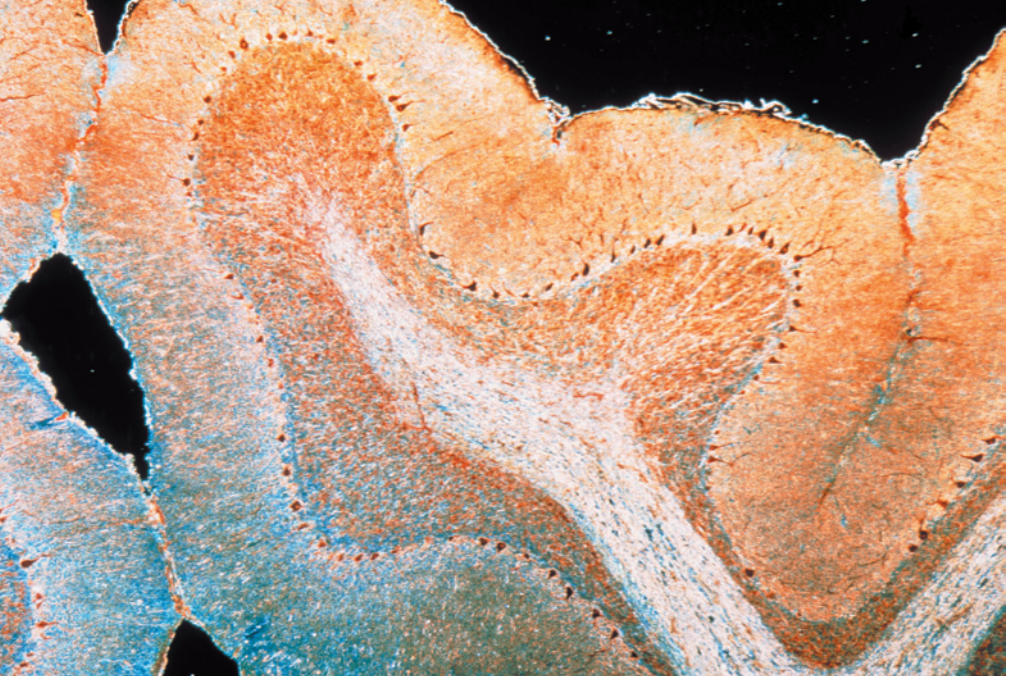
Nakashima, who will be assisting IPBES's work in this area, points to the use of fire by Australian Aborigines to manage tropical savannas as an example. Scientists had once thought that fire was damaging to the environment, but they now realize that it helps key species to propagate. The ability to incorporate such knowledge "will be be a key measure of the platform's success", says Larigauderie.

But Thomas Brooks, head of science at the IUCN, is worried that the platform's budget — US$43.5 million over five years, half of which has been pledged so far — is "much too small". He fears that the lion's share will go towards assessments, leaving little for activities such as boosting monitoring capabilities in poor countries. "There needs to be a more balanced allocation," he says.

Brooks also urges IPBES to develop guidelines to enable partners such as the IUCN to offer expertise and in-kind contributions, such as providing staff, to the assessments.

Larigauderie says that about 40% of funds — roughly $1.2 million in 2014 — will go to boosting monitoring capabilities, and adds that guidelines for partnerships, which should not preclude in-kind contributions, are likely to be finalized at the IPBES meeting in January. The biodiversity effort, she says, is open for business. ■

© 2014 Macmillan Publishers Limited. All rights reserved



EYE OF SCIENCE/SPL

Computer programs are being developed to stitch together images of brain slices to map neural circuits.

NEUROSCIENCE

Brain wave hits California

State creates programme to boost neuroscience innovation.

BY HELEN SHEN

As US science agencies firm up plans for a national ten-year neuroscience initiative, California is launching an ambitious project of its own. On 20 June, governor Jerry Brown signed into law a state budget that allocates US$2 million to establish the California Blueprint for Research to Advance Innovations in Neuroscience (Cal-BRAIN) project.

Cal-BRAIN is the first state-wide programme to piggyback on the national Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative announced by US President Barack Obama in April 2013 (see *Nature* **503,** 26–28; 2013). The national project is backed this year by $110 million in public funding from the National Institutes of Health (NIH), the Defense Advanced Research Projects Agency (DARPA) and the National Science Foundation (NSF).

California researchers and lawmakers hope that the state's relatively modest one-time outlay will pave the way for a larger multiyear endeavour that gives its scientists an edge in securing grants from the national initiative. "It's a drop in the bucket, but it's an important start," says Zack Lynch, executive director of the Neurotechnology Industry Organization, an advocacy group in San Francisco, California.

Cal-BRAIN sets itself apart from the national effort by explicitly seeking industry involvement. The proposal emphasizes the potential economic benefits of neuroscience research and calls for the formation of a programme to facilitate the translation of any discoveries into commercial applications.

"It's been difficult for the national initiative, particularly the NIH, to figure out exactly how to partner with the private sector," says Story Landis, director of the National Institute of Neurological Disorders and Stroke in Bethesda, Maryland. "In California, there's a really good history of these partnerships."

The NIH has so far committed $40 million towards the ten-year national programme, with DARPA contributing $50 million and the NSF adding another $20 million. Many neuroscientists say that this budget is much too small; on 6 June, scientific advisers to the NIH recommended that the agency increase its annual share to ten times the present value by 2020 and to $500 million a year for the five years after that.

The national initiative also involves private institutions, including two in California: the Salk Institute for Biological Studies in La Jolla and the philanthropic Kavli Foundation in Oxnard have committed $68 million over the project's lifetime.

"It just makes natural sense that California would be involved in a very serious way with the federal effort," says Democratic state senator Ellen Corbett, who introduced the legal provisions to create Cal-BRAIN.

Google, of Mountain View, California, is particularly interested in working with scientists on brain-inspired computing and on brain-mapping algorithms. The company is already collaborating with the Allen Institute for Brain Science in Seattle, Washington, to design better computer programs for building connectomes — three-dimensional maps of neural circuits — from images of brain slices.

"We would love to one day provide a service to universities in California and perhaps even companies, where they would give us their data, we would process it for them, and provide them with their connectome," says Thomas Dean, a research scientist at Google.

Many researchers have drawn parallels between Cal-BRAIN and the California Institute for Regenerative Medicine (CIRM), a stem-cell agency in San Francisco. But the analogy is not perfect: CIRM, established in 2004 and intended to support projects then ineligible for federal grants, has enjoyed the security of a $3-billion commitment from the state over ten years. Cal-BRAIN's $2 million in guaranteed funding is likely to disappear quickly, so its organizers will have to work fast to set up the programme and secure further support.

"Uncertainty of funding, that is the biggest difference between this and CIRM," says former CIRM science director Arlene Chiu.

The University of California has been tapped to coordinate Cal-BRAIN and to develop a governing structure with help from other universities, laboratories and businesses. The governing body would then craft a scientific plan and establish a competitive grants programme. Brooke Converse, a spokeswoman for the university, says that it is too early to reveal who might be involved in the planning process, how much of the budget would be spent on administration and how much would be left for grants.

"I don't think it's going to fund a lot of research, but it may be a way of starting to get organized," says Ralph Greenspan, associate director of the Kavli Institute for Brain and Mind at the University of California, San Diego.

Other states are watching with interest. In New York, scientists have been quietly meeting with lawmakers to discuss a similar initiative. Rafael Yuste, a neuroscientist at Columbia University in New York City whose ideas helped to spawn the US BRAIN Initiative, says that he and other researchers in the state are pushing for a neurotechnology research incubator in New York. "I'm sure that there will be other states that will feel that they need to do this as well," says Landis. ■

CORRECTION

The Editorial 'Summer skills' (*Nature* **510,** 312; 2014) said that only one Romanian student was selected for the summer school this year. In fact, two students from Romanian institutions made it through.

© 2014 Macmillan Publishers Limited. All rights reserved

Close collaborators

Romance often sparks between colleagues, and scientists are no different. *Nature* profiles four super-couples who have combined love and the lab.

BY KERRI SMITH

When physicists Claudia Felser and Stuart Parkin were introduced at a conference on applied magnetics, they felt an immediate attraction. But then, standing outside the Amsterdam conference centre, they started talking shop. It did not go well.

Parkin was interested in finding materials he could use to make miniature data-storage devices. Felser espoused the benefits of her pet topic: Heusler compounds, alloys with modifiable magnetic properties. "But he was not interested!" she laughs. Parkin thought that the compounds sounded as though they would be too difficult to interface with other materials. "So this was not a successful introduction," Felser says.

But the two kept in touch. And as Felser shared her growing knowledge about the semiconductor and quantum properties of Heusler compounds, Parkin grew more curious about the molecules — and about Felser. At the end of 2009, she decided to take a sabbatical from Johannes Gutenberg University in Mainz, Germany, to work at the headquarters of IBM in San Jose, California, where Parkin worked. "I invited her to stay with me," Parkin says. They were a couple from then on. "So this was more or less how it started and we're still working together," he says.

Felser and Parkin are one of thousands of couples who met through science. According to a 2010 survey by the US National Science Foundation, just over one-quarter of married people with doctorates had a spouse working in science or engineering[1]. Such partnerships are on the rise: in 1993, the proportion was one-fifth. More and more institutions are hiring couples. A 2008 survey[2] of around 9,000 US researchers found that the proportion of hires that went to couples rose from 3% in the 1970s to 13% in the 2000s. And data from the online dating service PlentyOfFish reveal that users with a graduate degree are three times more likely than the average user to form a couple with someone with a similar level of education.

Collaboration is key to the scientific process, but when collaborators are romantic partners, that relationship offers some unique advantages — a deep understanding of each other's personality and motivations — as well as the risk that work will dominate conversation at the dinner table. Here *Nature* talks to four couples about how they have managed to blend their science and lives.



MARK BAJOHRS/MAINZ

Claudia Felser and Stuart Parkin in the Gobi Desert in 2011.

MATERIALS AND AIR MILES

After Felser returned from her sabbatical, she and Parkin began racking up air miles. And Parkin's practical attitude rubbed off on Felser. "As a chemist you want to understand bonding, you want to find new synthesis methods. But you don't think deeply about applications," she says. Now, she started to also consider the material's cost and stability. As a result, companies lined up to work with her. "You really learn to think differently," she says. In 2011, the couple published a paper[3] on Heusler compounds and their potential in spintronics, a discipline that makes use of electrical fields to manipulate the spin of electrons.

Over the past few years Felser and Parkin have managed to spend up to one-quarter of their time together. Conferences and meetings became fruitful ways of meeting up. "As soon as people recognized that were are a couple, they started to invite us together to conferences. It was very good," Felser says.

Felser's employers — she is now director of the Max Planck Institute for Chemical Physics in Dresden — even realized that they might be able to persuade Parkin to accept a position in Germany. After many years on different continents, he is finally making arrangements to move, having been appointed director of the Max Planck Institute for Microstructure Physics in Halle. In April, he was awarded the Finnish Academy of Technology's Millennium Technology Prize, and plans to put part of the €1 million (US$1.4 million) in prize money towards building a house by the river in Halle. They plan to marry in December — on Stuart's birthday, "so I won't forget", he says. It will be their first place together. "Lufthansa and United will be very unhappy," Parkin says.

© 2014 Macmillan Publishers Limited. All rights reserved

NEURONAL CONNECTION

Lily and Yuh-Nung Jan have made their career studying cell division. But they themselves are inseparable. They start their sentences with 'we' or 'our'. Even their labs are joined. They met in 1967 in their native Taiwan when both were studying physics. Yuh-Nung had just got his bachelor's degree and his class was taking a celebratory hiking trip in the mountains. Joining them was a student from the class below: Lily. She had jumped ahead a year, catching up with Yuh-Nung, and was applying for graduate school, too. "I have a theory that quite a lot of her classmates were intimidated by her," says Yuh-Nung. "But I didn't know better."

Both got places studying physics at the California Institute of Technology (Caltech) in Pasadena. They were an item, but spent their first three years in separate dorms. Not long after they started work, a physicist-turned-biologist came to their department to give a seminar, and made them rethink their career choice. "Back in Taiwan we were not exposed to modern biology," says Yuh-Nung. "At Caltech, that was our first time. I guess it was good timing because biology was getting really interesting." Besides, he adds, tongue in cheek, they were over the hill as physicists. "All the great ones do something really important very early in their career, in their twenties, and we'd already reached that age."

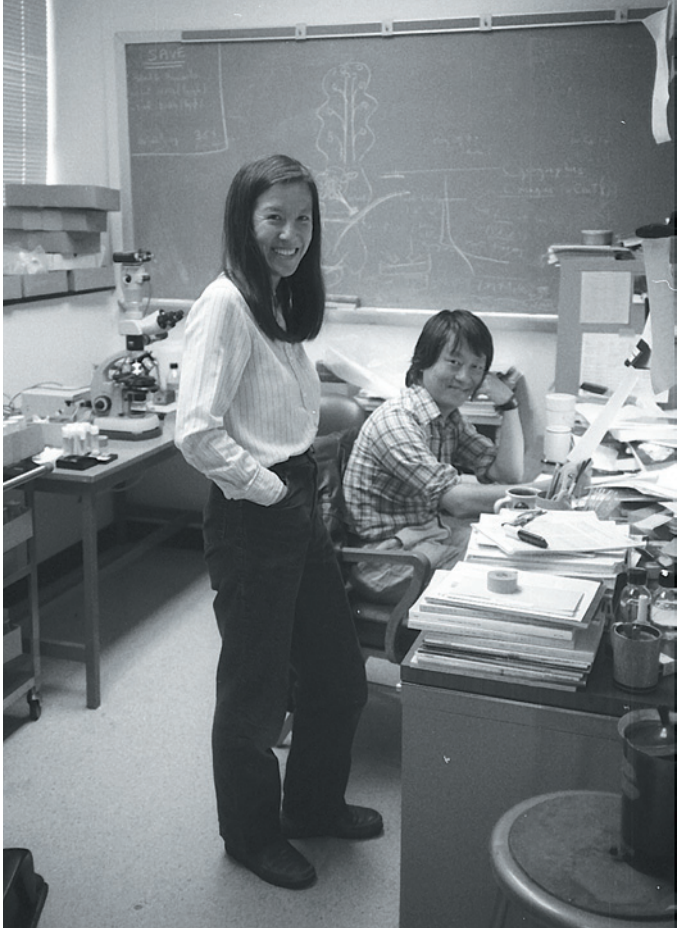
In a month they had made the switch to cell biology, and after pursuing separate thesis projects, began to collaborate. In 1971, they married. It was a very low-key ceremony at the Los Angeles courthouse — costing just US$6 for the licence and parking — and they celebrated by going camping and hiking in Yosemite National Park.

In 1979, they moved to the University of California, San Francisco. And having spent several years working in the same labs on similar projects, it was natural for them to run a lab together.

There were cases of too many cooks spoiling the broth. "In the very beginning, we both would sit with a postdoc or a student, and that certainly didn't work because no two people have the same idea," says Lily. "It very quickly evolved into an argument. The student was just looking back and forth."

Their interests overlapped heavily, but were sufficiently different that it made sense for Lily and Yuh-Nung to take the lead on different strands of the same problem: how brain cells divide. They now run adjoining labs, supervise 29 researchers and consistently produce publications in top journals. Lily focuses on ion channels and Yuh-Nung on cell morphology and, increasingly, function.

The Jans feel that being a couple gives them benefits over and above non-romantic collaborators. "It's not the sum of two parts, it's much better than that," says Lily. She puts their success down to "very consistent long-term camaraderie". And the pairing is certainly convenient. "Because whenever you think of something," she says, "it could be at home or at work, you can more easily discuss the questions." Yuh-Nung adds, "We've been together more than 40 years and I feel very lucky to have her as a partner."



Yuh-Nung and Lily Jan in their shared office in the early 1980s.

Their relationship seems to have served as a template for their colleagues. "There were some romances that started in the lab," says Lily. "More than one," Yuh-Nung says. "There have been kids born during the time their parents were in our lab, he says. "We lost track, but at some point we're going to put together an album."



Ruth Mace, Mark Pagel and their son Thomas in 1994.

FAMILY TREES

Few researchers can claim to have established a new field of science — let alone to have done so with their spouse. But that is exactly what evolutionary biologist Mark Pagel and anthropologist Ruth Mace did. They are the pioneers of using phylogenies — evolutionary trees — in anthropology, seeking to explain human cultures and behaviour as if they were evolving species.

When they first met, in the zoology department at the University of Oxford, UK, in the late 1980s, their work had little overlap. Mace was working on animal biology and Pagel was developing ways to analyse species relatedness. Both were heavily influenced by the evolutionary biology they were studying. British evolutionists, particularly, were known for their views on the power of adaptation and natural selection to explain behaviour. "We're both out of that church," says Mace. They first met at the department morning ▶

© 2014 Macmillan Publishers Limited. All rights reserved

break, which provided ample time to discuss their ideas. "Those were the days," Mace recalls. "The entire department would have massive amounts of coffee for an hour."

Several years later, Pagel and Mace co-authored a paper[4] that used phylogenetic methods to analyse human cultures, and argued that just as zoologists use genetics to look at species evolution, anthropologists could use languages to study human cultural evolution. The same year, their first son was born, adding a twig to their own tree of life.

Although they still collaborate on articles and research projects — Mace estimates that about 10% of their work is joint — they retain separate research identities. Both have academic interests outside their phylogeny work. But working in overlapping domains can lead to some awkward situations — especially because they have different surnames. Sometimes, one is asked to review the other's paper or a competing grant application — offers that they refuse with an explanation of the conflict of interest. "Being in the midst of two fields that have a history of 'robust' discussion, for want of a better word", Pagel is grateful to have someone who is on the same side.



COURTESY OF BORIS WORM AND HEIKE LOTZE

Boris Worm and Heike Lotze at Maasholm field station in 1998.

DREAM TEAM

Sometimes during his graduate work in marine ecology, Boris Worm would solve problems in his sleep. On waking, he would tell his partner, Heike Lotze, about his dream. A marine ecologist herself, Lotze served as a sleepy sounding board. "You know how you forget dreams in the morning. But if there's somebody next to you, you can tell them right away," says Worm.

The ecologists believe that their relationship has helped them to shape the early phases of their work in ways that would not be possible in a non-romantic collaboration. "We can share ideas as they emerge, very raw, very unfinished and some of it not useful but still interesting," Worm says. "I often have creative, intuitive ideas," adds Lotze. "Then I feel like I'm handing this raw thing over to Boris and he shapes it a bit."

Worm and Lotze met in the mid-1990s during their graduate study in Germany. Their fields overlapped, but they were pursuing different directions. Lotze was interested in the human influence on the sea and was studying nutrient pollution, thought to be the cause of algal blooms. She puts her practical mindset down to the fact that she was brought up on a farm. By looking after calves and baling hay, she routinely faced the connection between humans and the natural world, and how one changes the other. Worm's background is more analytical; his outlook more theoretical. The son of a psychologist and a professor of education, he grew up thinking a lot about relationships and communities. His PhD was on species interactions, particularly predation in ecosystems. "Heike's perspective grounded my ideas and gave them wheels, and maybe I have provided some wider context for the questions she was asking," he says.

They worked together throughout their PhDs — even using the same study site, Maasholm field station on the Baltic Sea. "There was an old rocket-launching station from the cold war, and part had been bought by our institution as a field site. We had the whole place to ourselves," says Worm. Because their experiments were often closely related, they had to do a little untangling before submitting their work for publication. "We had to sit down and say, OK, this is what I will publish and this is what you will publish," Lotze says.

They published their first big paper[5] together in 2002 — a grand synthesis of their PhD projects on the cumulative effects of various influences on marine ecosystems — and continue to publish together often. Perhaps their most controversial paper, produced as part of a large team in 2006, was a gloomy forecast of global fish stocks[6]. Worm and Lotze were dismayed by how much the media focused on "the end of seafood"; they had wanted to emphasize the rippling effects on species that are not harvested by humans. "The focus of the paper was different to what came out in the media," Lotze remembers. The phone rang constantly, and each found it helpful to have the other for support. "You understand what the other person is going through," says Lotze. "I'm a much more shy person, so for me to deal with those media — it was a storm, really. Boris was more riding the wave."

They are aware that their different personalities sometimes lead to Worm getting more attention than Lotze for their joint work. "Boris was more often the first spokesperson about our ideas," Lotze says. "For a while I kept a bit more in the background. People saw Boris more than me." But Lotze eventually started to step forward. "I didn't like being in the shadow, I had to fight that and get out of my shell," she says. They are occasionally told that they should differentiate their work, and have made a conscious effort not to co-author all their publications.

But last year, the couple won their first joint honour, the Peter Benchley Ocean Award for Excellence in Science. "It's not very often the connection gets recognized officially," Worm says. "It felt really wonderful to have that highlighted."

Official recognition is one thing, but for Lotze and Worm the greatest benefits of collaborating with a partner are less tangible. A romantic partner knows how to motivate, how to comfort when a grant proposal doesn't go your way and how to rein in the loopiest ideas. As Lotze says, "Your partner is your best critic." ■

Kerri Smith *writes for* Nature *in London.*

1. National Science Foundation *Survey of Doctorate Recipients* (NSF, 2010).
2. Schiebinger, L., Davies Henderson, A. & Gilmartin, S. K. *Dual-Career Academic Couples: What Universities Need to Know* (Stanford Univ., 2008).
3. Graf, T., Felser, C. & Parkin, S. S. P. *Progr. Solid State Chem.* **39,** 1–50 (2011).
4. Mace, R., Das Gupta, B. K., Bowen, J. R. & Pagel, M. *Curr. Anthropol.* **35,** 549–564 (1994).
5. Worm, B., Lotze, H. K., Hillebrand, H. & Sommer, U. *Nature* **417,** 848–851 (2002).
6. Worm, B. *et al. Science* **314,** 787–790 (2006).

© 2014 Macmillan Publishers Limited. All rights reserved



VITAMINS ON TRIAL

After decades of study, researchers still can't agree on whether nutritional supplements actually improve health.

BY MELINDA WENNER MOYER

ILLUSTRATION BY GARY NEILL

In 1911, Polish biochemist Casimir Funk discovered what was behind a then-mysterious neurological condition known as beriberi, common in regions where people's main source of calories came from de-husked, or 'polished', rice. He fed a group of ill pigeons a substance he had isolated from rice polishings, and within 12 hours, they had recovered. Funk went on to propose[1] that a handful of puzzling ailments including beriberi and scurvy arose because of deficiencies in nutrients like the one he had found in the rice husks. He considered these chemicals vital amines, which he shortened to "vitamines".

Although many embraced the idea that vitamins could prevent or reverse certain illnesses, the medical establishment railed against it: Funk's colleagues at the Lister Institute of Preventive Medicine in London questioned his theory and tried to ban him from using the term vitamine in his papers, and a 1917 editorial in the *Journal of the American Medical Association* noted that although "the expression 'deficiency disease' has become popular", the concept is a "vague explanation that is readily accepted by the uncritical"[2].

Today, nobody doubts that vitamin B1 can prevent beriberi or that vitamin C prevents scurvy. But scientific opinion about the use of vitamin supplements by millions of seemingly healthy people has never been more divided.

An editorial published in the *Annals of Internal Medicine*[3] last year offers a striking case in point. In it, researchers at Johns Hopkins University in Baltimore, Maryland, and other institutions proclaimed with certainty that the US public should "stop wasting money" on vitamin supplements. They argued that research has found no benefits, in part because most people in industrialized nations are well-nourished. Within months a counterattack ensued, headed by huge names in nutrition science and biochemistry, including Bruce Ames at the Children's Hospital Oakland Research Institute in California and Walter Willett at Harvard University in Cambridge, Massachusetts, who argued that vitamin deficiencies are, in fact, widespread in the United States and

© 2014 Macmillan Publishers Limited. All rights reserved

that supplements can help to close nutritional gaps[4].

Meir Stampfer, an epidemiologist at Harvard, considers the anti-vitamin editorial "garbage". "I just felt sadness that such a poorly done paper would be published in a prominent journal and cause so much confusion," he says.

The argument raises important questions about the quality and relevance of more than a century of studies. "There will always be two polar sides to this argument, and the main reason for that is we don't know the answers — we don't have evidence one way or the other," says Paul Coates, who directs the Office of Dietary Supplements at the US National Institutes of Health (NIH) in Bethesda, Maryland.

Pooled together, evidence from double-blinded, placebo-controlled clinical trials suggests that hardly any nutrient supplements have a consistent health effect on people in developed countries. But many argue that the null findings reflect research deficiencies, including poor study design, inappropriate mixing of different kinds of data and misunderstandings about how much of a nutrient is enough. "The tools we've had in the past have been so crude — it's like we've been looking through a dirty window with the curtains closed," says Susan Mayne, chair of the department of chronic disease epidemiology at the Yale School of Public Health in New Haven, Connecticut, and a member of the Institute of Medicine's Food and Nutrition Board, which establishes US nutrition guidelines such as the Dietary Reference Intakes for vitamins and minerals.

Although some scientists say that researchers can glean important insights from existing nutrition data, others, such as Robert Heaney, an endocrinologist at Creighton University in Omaha, Nebraska, argue that most existing studies are fatally flawed and that the whole enterprise needs a methodological overhaul.

"The trials we have in nutrition aren't answering the right questions, so they're not appropriate," says Connie Weaver, head of the department of nutrition science at Purdue University in West Lafayette, Indiana, and a member of the Food and Nutrition Board. "What we're using now is pretty bad science."

THE SWEET SPOT

The market for vitamins and supplements has been estimated at US$68 billion worldwide — and multivitamins are by far the most popular. Most people take them not to treat diagnosed deficiencies, but to improve or maintain "overall health", as reported[5] last year by the Office of Dietary Supplements. Clinical deficiencies such as scurvy are rare in industrialized nations, but some research suggests that many people are at least mildly deficient in certain nutrients. In 2011, an analysis of data from the US Centers for Disease Control and Prevention's National Health and Nutrition Examination Survey (NHANES) reported[6] that more than one-quarter of the US population are not getting enough of vitamins A, C, D and E, calcium or magnesium. Ninety-seven per cent fail to get enough potassium. The current Dietary Guidelines for Americans, released in 2010, warn that the intake of potassium, dietary fibre, calcium and vitamin D among the general public is "low enough to be of public health concern". Scientists argue over just how important these seemingly widespread subclinical deficiencies are. Johns Hopkins epidemiologist Pete Miller dismisses them as nonexistent. He argues that even with government-recommended nutrient levels, "the threshold for what defines deficiency is probably incorrect".

Surveys used to determine what kind of food people eat, and how much, are notoriously unreliable, for example. In 2013, a *PLoS ONE* study[7] found that energy-intake data from the 39 years of NHANES studies were "not physiologically plausible" for a majority of participants because of systematic under-reporting.

Still for some nutrients, Mayne says, it is almost certain that some people are "really not getting enough". Moreover, research suggests that individuals with lower-than-average intakes but clinically 'normal' nutrient statuses could still benefit from supplements. Researchers at the Harvard School of Public Health recruited 672 health professionals with histories of benign colorectal tumours, a risk factor for colorectal cancer, to see if folic acid helped to reduce tumour recurrence. Half of the participants took 1 milligram of folic acid a day for between 3 and 6.5 years, and the other half took a placebo. The supplements had no effect when everyone was analysed together, but among the people with the lowest folic acid intake at the start of the trial, those who took supplements had a reduced risk of recurrence[8].

"The tools we've had in the past have been so crude — it's like we've been looking through a dirty window with the curtains closed."

On the other side of the coin, several large trials suggest that over-consuming nutrients could be dangerous. The Alpha-Tocopherol Beta-Carotene Cancer Prevention Trial set out to see whether smokers would benefit from certain supplements. It turned out that those who took 20 mg of the vitamin A precursor β-carotene a day — 3 times the US recommended daily allowance for vitamin A — for 5–8 years were, in fact, 18% more likely to develop lung cancer than those taking a placebo[9]. A potential explanation is that the breakdown products of β-carotene can, at high doses, cause cell proliferation.

These results illustrate one of the many complexities of nutrient metabolism. Nutrition scientists now recognize that risk curves are J- or U-shaped: nutrients have beneficial effects at low doses and toxic effects at high doses. The magnitude of the response differs, too, depending on where individuals start on the curve — their baseline status.

Yet this is often ignored. Many of the studies included in a 2009 systematic review[10] commissioned by the US government to inform the development of guidelines for vitamin D intake, for instance, did not include information on baseline status, so the review probably lumped together individuals who responded differently; perhaps unsurprisingly, it concluded that findings on vitamin D were "inconsistent" (see *Nature* **475,** 23–25; 2011). "These are common-sense design issues that need to be put on the table when you're thinking about a study, and they haven't been," says Heaney.

MASKED EFFECTS

Nutrient intake in control groups is also important but often overlooked. As part of the NIH's Women's Health Initiative (WHI), researchers tested the effects of daily doses of 1,000 mg of calcium a day — along with vitamin D — on women's fracture risks[11]. Although NHANES data at the time suggested that the average calcium intake for postmenopausal women was about 600 mg a day, the trial investigators learned after randomization that women in the control group were actually consuming more than 1,000 mg a day.

The trial found no statistically significant difference in fracture risk between the intervention and control groups, but "the study design was such that it couldn't show anything", Heaney says. Even so, "systematic reviews continue to include the Women's Health Initiative as a 'negative' study".

The trial highlights two other potential confounding factors. The first is that study participants are typically more health-conscious than other individuals. Although WHI participants on average consumed relatively high levels of calcium, nearly 75% of US women between the ages of 31 and 50 fail to get the recommended 1,000 mg of calcium per day from food. The second point is that compliance with instructions is often low — only 59% of participants were still taking at least 80% of their pills by the end. Those who do not stick to the prescribed treatments might differ from those who do in important ways, skewing outcomes.

Another important factor is genetic variability. "Every person has about 50,000 variations in their genes," says Steven Zeisel, director of the University of North Carolina Nutrition Research Institute in Chapel Hill. Any number

© 2014 Macmillan Publishers Limited. All rights reserved

DATA DEFICIENCIES

For almost every claim about vitamins, studies have found at least some negative evidence. But most of the negative findings have some confounding variables.

"CHOLINE SUPPLEMENTATION BOOSTS FETAL BRAIN DEVELOPMENT."

The negative finding: A double-blind, randomized controlled trial[15] involving 140 pregnant women found no enhancement to infant brain function.

The confounder: 44% of women have genetic variants that significantly increase their dietary choline requirement[16].

"CALCIUM AND VITAMIN D SUPPLEMENTATION REDUCES THE RISK OF BONE FRACTURES."

The negative finding: The Women's Health Initiative calcium and vitamin D trial[11] found no effect of supplementation on risk of hip fractures.

The confounder: Researchers had estimated that the control group was consuming less than 600 mg calcium per day, when actual intake was more than 1,000 mg.

"β-CAROTENE REDUCES THE RISK OF COLORECTAL ADENOMA, A PRECURSOR TO COLORECTAL CANCER."

The negative finding: A 1994 clinical study[17] assigning 864 patients to different treatment groups that included β-carotene supplementation found no evidence of benefit, and some evidence of harm.

The confounder: Among non-smokers and non-drinkers there was a significant decrease in the risk of adenocarcinoma, but smokers and drinkers had an increased risk[18].

"ANTIOXIDANT SUPPLEMENTATION REDUCES CANCER AND MORTALITY RISK."

The negative finding: A meta-analysis[19] of 21 clinical trials, covering a pooled sample of around 91,000 people and 8,800 deaths, found no evidence of an effect on mortality risk.

The confounder: The analysis did not stratify results by sex. Men may have been more likely than women to benefit from antioxidants[20].

of them could be important in metabolism. Yet "very few geneticists are collecting diet information, and very few diet people collect genetic information". Zeisel's work has uncovered, for example, that 44% of women have gene variants that significantly increase their dietary requirements for the nutrient choline. It is perhaps no wonder that trial results have been inconsistent — and that reviews often report null findings (see 'Data deficiencies'). Plus, the effects of nutrition interventions are probably subtle: whereas drug trials compare exposure with no exposure, nutrition trials compare higher and lower exposures, because everyone eats and consumes some nutrients. Subtle differences may be hard to detect and have long latency periods. These limitations and considerations add up "in a way that causes trials to be heavily stacked against showing any benefit", says biochemist Balz Frei, director of the Linus Pauling Institute at Oregon State University in Corvallis.

A CLEARER VIEW

So how can scientists design studies to arrive at nutritional truths? In a paper[12] published in *Nutrition Reviews* in January, Heaney proposes guidelines. He argues that, first and foremost, scientists need to consider the dose–response curve. It is absolutely crucial, he says, to measure the baseline nutritional status of trial participants and track changes over time. He also suggests that investigators use participants with similar baseline intakes. That could limit who the results apply to, but it would make data clearer.

Measuring nutrient status presents additional challenges. Take calcium: the body carefully regulates blood levels, keeping them constant by pulling the mineral from bone when intake is low. Heaney says that it might be possible to measure how much is being consumed by looking at other biomarkers, such as parathyroid hormone, which triggers the removal of calcium from bone, but such tests can be expensive.

Researchers must also come up with accurate ways to assess food and nutrient intake during trials. This requires better estimates of the nutrients in foods. Frei says that the US Department of Agriculture's National Nutrient Database for Standard Reference, considered the country's authoritative source of food-composition data, overestimates the amounts of vitamin A in foods because the standard units used do not account for nutrient bioavailability. And it misses some other sources of vitamins in food. It does not, for example, assess levels of 25-hydroxycholecalciferol, a form of vitamin D found in animal products.

The good news is that scientists have developed technologies to improve measurements of both nutrient status and intake. Mayne and her colleagues, for instance, have created a skin-spectroscopy method for assessing levels of carotenoids[13]. "You can do it for zero cost, essentially — it's a 30-second scan on someone's skin that gives a readout of their nutrient status," she says. And researchers at Purdue are building a smartphone app that tracks food intake. Users take a picture of a plate of food, and the application estimates and records its nutritional components. These estimates have been shown to be more accurate than self-reports[14].

Researchers also need to do better at considering and accounting for confounding factors. And reviews should be designed to answer specific questions about specific nutrient doses in specific populations, says Heaney — basically, they should include only those trials that "are appropriate for collapsing into a single data set". The Office of Dietary Supplements has sponsored a series of technical reports to investigate the challenges associated with evidence-based reviews in nutrition; six have been published.

So, back to the original question. Are supplements useless? The current state of research offers only an equivocal half-answer: 'maybe yes' for some individuals, nutrients and doses, and 'maybe no' for others. "Nutrition is complex, and I don't think we're necessarily going to find one formula that works for everybody," says Mayne. But new tools in development could "really change the way we look at this". The big question is whether, once scientists have all the pieces they need, they can put them together to create a clear and cohesive picture. ■

Melinda Wenner Moyer *is a freelance writer in Cold Spring, New York.*

1. Funk, C. *J. State Med.* **20,** 341–368 (1912).
2. *J. Am. Med. Assoc.* **24,** 2040–2041 (1917).
3. Guallar, E., Stranges, S., Mulrow, C., Appel, L. J. & Miller, E. R. III *Ann. Intern. Med.* **159,** 850–851 (2013).
4. Frei, B., Ames, B. N., Blumberg, J. B. & Willett, W. C. *Ann. Intern. Med.* **160,** 807 (2014).
5. Bailey, R. L., Gahche, J. J., Miller, P. E., Thomas, P. R. & Dwyer, J. T. *JAMA Intern. Med.* **173,** 355–361 (2013).
6. Fulgoni, V. L. III, Keast, D. R., Bailey, R. L. & Dwyer, J. *J. Nutr.* **141,** 1847–1854 (2011).
7. Archer, E., Hand, G. A. & Blair, S. N. *PLoS ONE* **8,** e76632 (2013).
8. Wu, K. *et al. Am. J. Clin. Nutr.* **90,** 1623–1631 (2009).
9. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group *N. Engl. J. Med.* **330,** 1029–1035 (1994).
10. Chung, M. *et al. Evid. Rep. Technol. Assess. (Full Rep.)* **183** 1–420 (2009).
11. Jackson, R. D. *et al. N. Engl. J. Med.* **354,** 669–683 (2006).
12. Heaney, R. P. *Nutr. Rev.* **72,** 48–54 (2014).
13. Mayne, S. T. *et al. Arch. Biochem. Biophys.* **539,** 163–170 (2013).
14. Lee, C. D. *et al. J. Diabetes Sci. Technol.* **6,** 428–434 (2012).
15. Cheatham, C. L. *et al. Am. J. Clin. Nutr.* **96,** 1465–1472 (2012).
16. Zeisel, S. H. *Ann. Nutr. Metab.* **60,** 19–25 (2012).
17. Greenberg, E. R. *et al. N. Engl. J. Med.* **331,** 141–147 (1994).
18. Baron, J. A. *et al. J. Natl Cancer Inst.* **95,** 717–722 (2003).
19. Macpherson, H., Pipingas, A. & Pase, M. P. *Am. J. Clin. Nutr.* **97,** 437–444 (2012).
20. Hercberg, S. *et al. Arch. Intern. Med.* **164,** 2335–2342 (2004).

© 2014 Macmillan Publishers Limited. All rights reserved

COMMENT

QUANTUM PHYSICS Where next, 50 years on from Bell's theorem? **p.467**

BIBLIOMETICS On the uses, misuses and future of impact measures **p.470**

SPACE Biography of Neil Armstrong is more missions than man **p.472**

RECESSION Evidence that Greek science has been hit hard by financial crash **p.473**

MICHAEL S. WILLIAMSON/*THE WASHINGTON POST* VIA GETTY



Oil extracted from fields in Alberta, Canada, would travel through the proposed Keystone XL pipeline.

Consider the global impacts of oil pipelines

Debates over oil-sands infrastructure obscure a broken policy process that overlooks broad climate, energy and environment issues, warn **Wendy J. Palen** and colleagues.

The debate over the development of oil sands in Alberta, Canada, is inflaming tensions in and between Canada and the United States.

In April, US President Barack Obama deferred a decision on the fate of the proposed Keystone XL oil pipeline, despite escalating pressure to approve it from Canadian Prime Minister Stephen Harper. The contentious pipeline would transport 830,000 barrels per day of partially refined bitumen from Alberta's oil sands, through the US Midwest, to Gulf Coast refineries. Harper is also facing a controversial domestic battle over his approval on 17 June of the Enbridge Northern Gateway pipeline, to connect Alberta with a port on British Columbia's remote Pacific coast.

But drama over the pipelines obscures a larger problem — a broken policy process. Both Canada and the United States treat oil-sands production, transportation, climate and environmental policies as separate issues, assessing each new proposal in isolation. A more coherent approach, one that evaluates all oil-sands projects in the context of broader, integrated energy and climate strategies, is sorely needed.

Although Keystone XL and Northern Gateway are among the first major North American projects to highlight flaws in oil-sands policies, more than a dozen other projects are on the drawing board. ▶

© 2014 Macmillan Publishers Limited. All rights reserved

Meanwhile, the US government is considering its first oil-sands leases on federal lands, as bitumen mining expands on state land in Utah's Uinta Basin.

As scientists spanning diverse disciplines, we urge North American leaders to take a step back: no new oil-sands projects should move forward unless developments are consistent with national and international commitments to reducing carbon pollution. Anything less demonstrates flawed policies and failed leadership. With such high stakes, our nations and the world cannot afford a series of ad hoc, fragmented decisions.

INCREMENTAL DECISIONS

Current public debate about oil-sands development focuses on individual pipeline decisions. Each is presented as an ultimatum — a binary choice between project approval and lost economic opportunity. This approach artificially restricts discussions to only a fraction of the consequences of oil development, such as short-term economic gains and job creation, and local impacts on human health and the environment. Lost is a broader conversation about national and international energy and economic strategies, and their trade-offs with environmental justice and conservation.

This pattern of incremental decisions creates the misguided idea that oil-sands expansion is inevitable. By restricting the range of choices, governments have allowed corporations to profit from one-off policy decisions, leading to a doubling of oil-sands production in Alberta in the past decade, with production forecast to double again to 3.9 million barrels per day in the coming decade[1]. The collective result of these decisions is unnecessarily high social, economic and environmental costs.

"Plans for the development of similar unconventional reserves are considered worldwide."

When judged in isolation, the costs, benefits and consequences of a particular oil-sands proposal may be deemed acceptable. But impacts mount with multiple projects. The cumulative effects of new mines, refineries, ports, pipelines, railways and a fleet of transoceanic supertankers are often at odds with provincial, state, federal or international laws protecting clean water, indigenous rights, biodiversity and commitments to control carbon emissions.

Oil-sands development in Alberta, for example, has irreversibly transformed more than 280 square kilometres of the boreal landscape by burning or degrading peatlands covering oil-sands deposits[2]. Such ecosystems represent long-term carbon sinks that require thousands of years to develop. The development has also elevated waterway concentrations of chemical contaminants such as polycyclic aromatic compounds that are toxic to fish and other aquatic organisms[3,4], and has been associated with a tenuous but troubling rise in rare cancers in downstream indigenous communities[5].



Major infrastructure such as pipelines requires decades of operation to recoup the initial investment, fostering expansion of oil-sands projects upstream and refineries and ports downstream. For example, the proposed US$5.4-billion Keystone XL pipeline would drive further oil-sands extraction by providing access to Gulf Coast refineries and profitable export markets (see 'Big decisions'). Such investments create a 'lock-in' that commits society to decades of environmental degradation, increased risk of contamination and spills, and unsustainable carbon pollution.

Oil-sands production has already caused dramatic increases in carbon pollution. The United States and Canada have committed to the same 2020 greenhouse-gas emissions target: a 17% decrease relative to 2005 levels. But Canada's agencies predict that it will miss its target by 122 million tonnes annually[6,7]. Although emissions in many sectors are falling, those from oil-sands production are predicted to triple from 2005 by 2020, from 34 million to 101 million tonnes.

SMART STEPS

Despite these predictions, public discussions around emissions from expanding oil-sands production are being muted. Since 2010, public hearings on proposed pipelines, including Northern Gateway, the Trans Mountain pipeline in British Columbia, and the Line 9B pipeline reversal in southern Ontario, have formally excluded testimony by experts or the public about carbon emissions and climate (see go.nature.com/mpx2sc).

We propose two steps to improve decisions about the development of oil sands. First, North American citizens and policy-makers must enact policies at national, state and provincial levels that acknowledge the global consequences of expanding oil-sands development. Legislated constraints on carbon pollution (such as a carbon tax or cap-and-trade) based on current climate science will help to ensure that the full social costs of carbon combustion are incorporated into investment decisions about energy and infrastructure. This will help companies and policy-makers to better judge trade-offs between investment in oil-sands projects, renewables and energy-conservation programmes, while catalysing innovation in low-carbon technologies.

Second, policy-makers need to adopt more transparent and comprehensive decision-making processes that incorporate trade-offs among conflicting objectives such as energy and economic development, environmental protection, human health and social justice. The decision sciences offer pathways, from problem identification to policy implementation, which can encompass a wide range of public values and address multiple drivers, linked effects and nested scales of cause and effect.

Decision-support tools are being developed for exploring how outcomes, priorities and trade-offs shift under different future energy scenarios. Possibilities might include the approval or rejection of pipeline proposals, more stringent low-carbon fuel standards, carbon taxes, or a spike or drop in global demand for Canadian oil[8]. Such tools can be used to identify thresholds where development should shift from one energy option to another, and evaluate which investments are most robust given environmental, social and economic policies and their effects on energy supply and demand[8]. The territorial government in the Canadian north is using these tools to identify energy options that protect the Arctic environment and developing economy, while meeting the needs of local communities[9].

In the absence of a global accord to reduce carbon emissions, the United States and Canada should agree to a suite of shared policies to guide development of both carbon-based and low-emission sources of energy over the coming decades. Such coordination might seem unlikely given the ideological gulf between the current US and Canadian administrations, but that divide will not persist indefinitely.

A binational carbon and energy strategy should align with existing continental trade accords, provide a clear road map for decisions about energy development — particularly for unconventional oil — and enhance North American competitiveness and leadership. It should specify priorities, expectations and principles whereby decisions on infrastructure projects, such as Keystone XL or Northern Gateway, are made in the context

© 2014 Macmillan Publishers Limited. All rights reserved

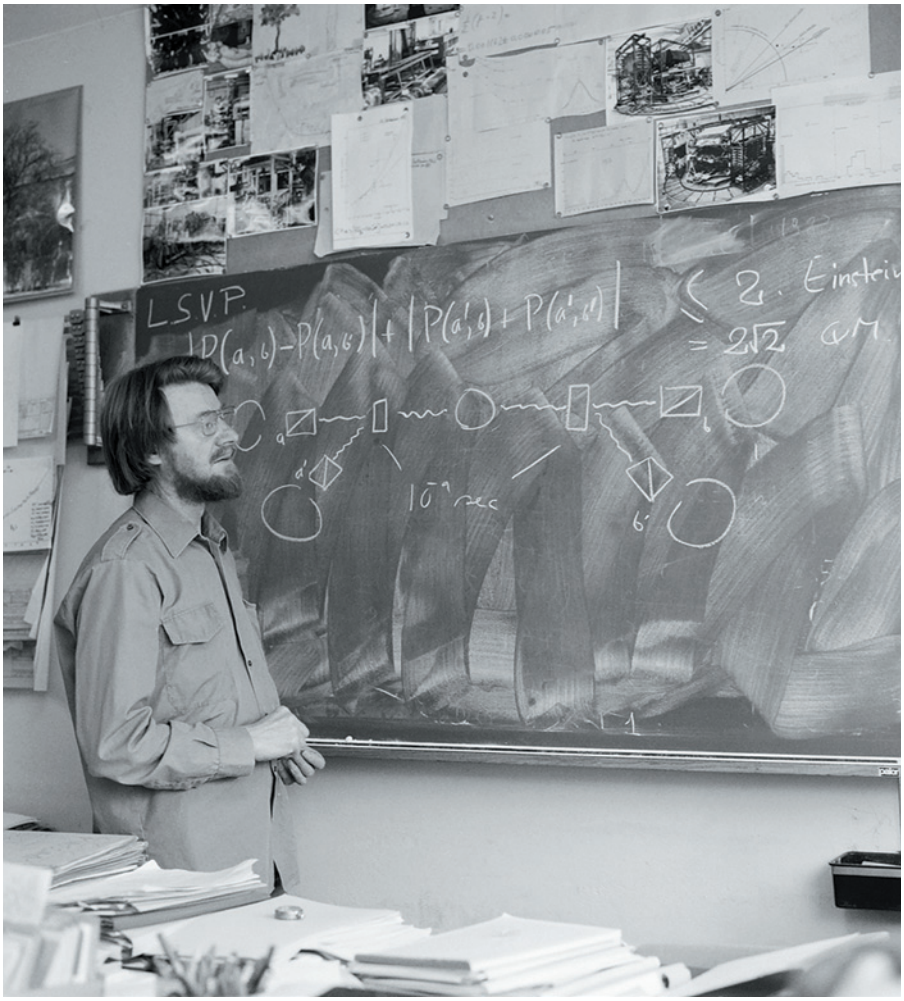
of an overarching commitment to limit carbon emissions. North America's energy challenges would then become a vehicle for beneficial economic coordination and integration rather than remaining a source of rancour and friction.

A key step is a moratorium on new oil-sands development and transportation projects until better policies and processes are in place. Reform is needed now: decisions made in North America will reverberate internationally, as plans for the development of similar unconventional reserves are considered worldwide.

With clearer policy, smarter decisions and stronger leadership, Canada and the United States can avoid the tyranny of incremental decisions — and the lasting economic and environmental damage that poorly conceived choices will cause. ■

Wendy J. Palen *is in the Department of Biological Sciences at Simon Fraser University, Burnaby, British Columbia, Canada.* **Thomas D. Sisk** *is in the School of Earth Sciences and Environmental Sustainability at Northern Arizona University, Flagstaff, Arizona, USA.* **Maureen E. Ryan** *is in the School of Resource and Environmental Management at Simon Fraser University.* **Joseph L. Árvai** *is in the Department of Geography at the University of Calgary, Alberta, Canada.* **Mark Jaccard** *is in the School of Resource and Environmental Management at Simon Fraser University.* **Anne K. Salomon** *is in the School of Resource and Environmental Management at Simon Fraser University.* **Thomas Homer-Dixon** *is in the Balsillie School of International Affairs at the University of Waterloo, Ontario, Canada.* **Ken P. Lertzman** *is in the School of Resource and Environmental Management at Simon Fraser University.*
e-mail: wpalen@sfu.ca

1. Canadian Association of Petroleum Producers. *2014 CAPP Crude Oil Forecast, Markets, & Transportation: Refinery Data* (CAPP, 2014).
2. Rooney, R. C., Bayley, S. E. & Schindler, D. W. *Proc. Natl Acad. Sci. USA* **109,** 4933–4937 (2012).
3. Kurek, J. *et al. Proc. Natl Acad. Sci. USA* **110,** 1761–1766 (2013).
4. Kelly, E. N. *et al. Proc. Natl Acad. Sci. USA* **106,** 22346–22351 (2009).
5. Chen, Y. *Cancer Incidence in Fort Chipewyan, Alberta 1995–2006* (Alberta Cancer Board, 2009).
6. Environment Canada. *Canada's Emission Trends* (2013).
7. Office of the Auditor General of Canada. *2012 Spring Report of the Commissioner of the Environment and Sustainable Development* Ch. 2 (2012).
8. Arvai, J., Gregory, R., Bessette, D. & Campbell-Arvai, V. *Iss. Sci. Technol.* **28,** 43–52 (2012).
9. Kenney, L., Bessette, D. & Arvai, J. *J. Environ. Plan. Mgmt* http://dx.doi.org/10.1080/09640568.2014.899205 (2014).



CERN

Physicist John Bell at CERN, Europe's particle-physics lab near Geneva, Switzerland, in 1982.

Bell's theorem still reverberates

Fifty years ago, John Bell made metaphysics testable, but quantum scientists still dispute the implications. **Howard Wiseman** proposes a way forward.

In 1964, Northern Irish physicist John Bell proved mathematically that certain quantum correlations, unlike all other correlations in the Universe, cannot arise from any local cause[1]. This theorem has become central to both metaphysics and quantum information science. But 50 years on, the experimental verifications of these quantum correlations still have 'loopholes', and scientists and philosophers still dispute exactly what the theorem states.

Quantum theory does not predict the outcomes of a single experiment, but rather the statistics of possible outcomes. For experiments on pairs of 'entangled' quantum particles, Bell realized that the predicted correlations between outcomes in two well-separated laboratories can be profoundly mysterious (see 'How entanglement makes the impossible possible'). Correlations of this sort, called Bell correlations, were verified experimentally more than 30 years ago (see, for example, ref. 2). As Bell proved in 1964, this leaves two options for the nature of ▶

© 2014 Macmillan Publishers Limited. All rights reserved

reality. The first is that reality is irreducibly random, meaning that there are no hidden variables that "determine the results of individual measurements"[1]. The second option is that reality is 'non-local', meaning that "the setting of one measuring device can influence the reading of another instrument, however remote"[1].

Most physicists are localists: they recognize the two options but choose the first, because hidden variables are, by definition, empirically inaccessible. Quantum information scientists embrace irreducible randomness as a resource for secure cryptography[3]. Other physicists and philosophers (the 'non-localist camp') dispute that there are two options, and insist that Bell's theorem mandates non-locality[4].

"Before investing too much angst or money, one wants to be sure that Bell correlations really exist."

Such views seem contradictory. But I believe that these two camps can be partially reconciled[5] by delving into what 'causation' means. Doing so reveals the depth of the real principles at stake, the challenges facing each camp, and the future priorities for closing the loopholes in experiments to observe Bell correlations.

FREE CHOICE

Many localists cite Albert Einstein's 1905 principle of relativistic causality as a reason to reject non-locality. This principle says that causal influences cannot propagate faster than light. That is, one event can cause another (later) event only if they are close enough in space that the 'effect' could have been reached by a beam of light from the 'cause'. But the non-locality option must involve faster-than-light causal influences, contrary to Einstein's principle, for measurements made far enough apart and close to simultaneous. Bell correlations under these circumstances have been observed many times since 1982 (ref. 2), using photons with entangled polarizations.

Although the two camps disagree on whether Bell experiments imply faster-than-light causal influences, neither think that these experiments allow faster-than-light communication. Faster-than-light communication has never been observed. Its impossibility follows from Einstein's principle of relativistic causality and the following axiom of causation: if an event is seen to depend statistically on a freely chosen action, then that action is a cause of that event.

For example, if a radio comes on when and only when I choose to flip a switch, then my action must cause the sound. Combining this with Einstein's principle (no faster-than-light causal influences) implies that if I freely choose the time to flip a switch on Earth, a radio on the Moon cannot be expected to come on at that exact time. There must be a delay of at least 1.3 seconds (the time it takes light to travel to the Moon).

It is wrong to argue (as some localists do) that the impossibility of faster-than-light communication rules out non-locality. Locality, as Bell introduced it in 1964, is a stronger concept than no faster-than-light communication. That is, nature could be non-local without allowing faster-than-light signalling.

For localists to derive locality from the principle of relativistic causality, they need a stronger version of the above axiom of causation: the phrase "is seen to depend on" must be replaced by "depends, in theory, on"[5]. The point is that it may not be possible to see the theoretical dependence if there are other, hidden variables on which the event also depends. This is the case in the versions of quantum theory favoured by the non-localist camp[4].

ANOTHER THEOREM

Bell himself was a non-localist, an opinion he first published in 1976 (ref. 6), after introducing a concept, "local causality", that is subtly different from the locality of the 1964 theorem. Deriving this from Einstein's principle requires an even stronger notion of causation: if two events are statistically correlated, then either one causes the other, or they have a common cause, which, when taken into account, eliminates the correlation.

Colloquially, this "principle of common cause" says that correlations have explanations. For example, if you and I never communicate, but one day we both become concerned about a looming war in Ruritania, then there must be a common cause for our thoughts (such as news reports).

In 1976, Bell proved that his new concept of local causality (based implicitly on the principle of common cause), was ruled out by Bell correlations[6]. In this 1976 theorem there was no second option, as there had been in the 1964 theorem, of giving up hidden variables. Nature violates local causality.

It is unfortunate that quantum scientists seldom distinguish the 1976 theorem from the 1964 theorem. It is doubly unfortunate that Bell sometimes used "locality" as shorthand for "local causality"[6], adding to the confusion. Non-localists maintain that the two theorems are the same, that locality is the same as local causality, and thus that hidden variables played no essential part in Bell's 1964 paper[4]. But, as I have shown[5], these claims do not hold up under careful analysis.

RECONCILING THE CAMPS

The contradictory claims by the two camps thus arise because they mean different things by 'Bell's theorem' and different things by 'local' (or 'non-local'). For localists, Bell's theorem is the 1964 one, and the preferred choice is to keep locality and forgo hidden variables. For non-localists, Bell's theorem is (or should be) the 1976 one, which leaves no choice but to forgo local causality.

But one can go further, by recalling that local causality rests on two principles: Einstein's principle of relativistic causality, and the principle of common cause. Thus Bell's 1976 theorem can be restated as: either causal influences are not limited to the speed of light, or events can be correlated for no reason.

This, I suggest, is the best way to reconcile the two camps. It enables them to agree on a single Bell's theorem, and what logical options it offers, even if they prefer different options.

Those who insist that correlations are explicable must conclude that causal influences can go faster than light. A challenge for these non-localists is: why does nature nevertheless conspire to prevent faster-than-light signalling?

Those who hold Einstein's principle to be inviolable (the localists) must conclude that some events are correlated for no reason. A challenge for them is: if correlations do not necessarily imply a cause, when should scientists look for causes, and why?

THE PATH FORWARD

Bell correlations can be seen as a problem, or an opportunity. They present us with a dilemma; each of the principles at stake (relativistic causality and common cause) underpins a vast mesh of scientific inference and intuition, and yet one must be forgone. But Bell correlations also present us with a marvellous information-technology resource: measurement outcomes that cannot possibly be known to anyone before they occur.

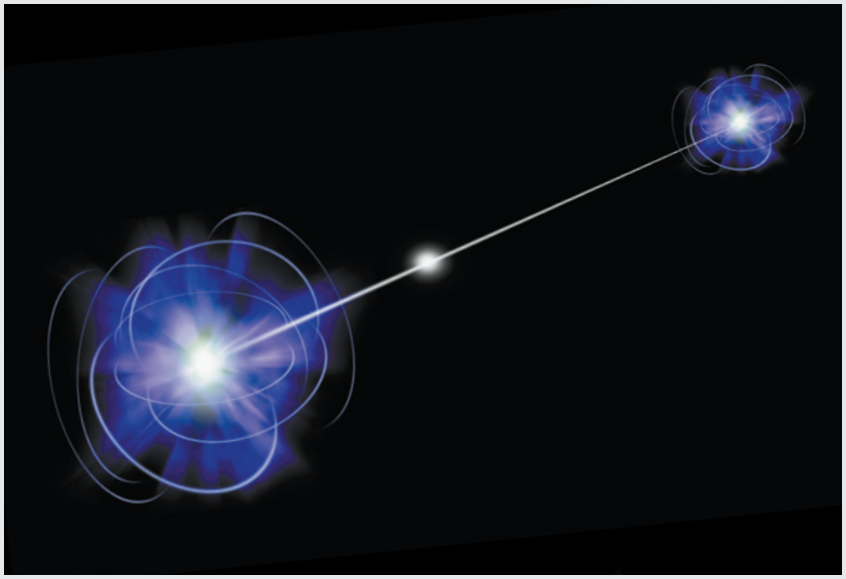
Before investing too much angst or money, one wants to be sure that Bell correlations really exist. As of now, there are no loophole-free Bell experiments. Experiments in 1982 by a team led by French physicist Alain Aspect[2], using well-separated detectors with settings changed just before the photons were detected, suffered from an 'efficiency loophole' in that most of the photons were not detected. This allows the experimental correlations to be reproduced by (admittedly, very contrived) local hidden variable theories.

In 2013, this loophole was closed in photon-pair experiments using high-efficiency detectors[7,8]. But they lacked large separations and fast switching of the settings, opening the 'separation loophole': information about the detector setting for one photon could have propagated, at light speed, to the other detector, and affected its outcome.

There are several groups worldwide racing to do the first Bell experiment with large separation, efficient detection

© 2014 Macmillan Publishers Limited. All rights reserved

VICTOR DE SCHWANBERG/SPL



Quantum entanglement can link the quantum states of particles even when they are separated by long distances (artist's impression).

BELL CORRELATIONS

How entanglement makes the impossible possible

Consider an impossible square — a square divided into nine smaller squares, each containing a 0 or a 1, such that the number of 1s in every column is even, and the number of 1s in every row is odd. Why is this impossible? Because the total number of 1s, from the column-rule, is even + even + even = even, but the total number of 1s from the row-rule is odd + odd + odd = odd.

Two shady characters, Rowan and Colin, approach you, claiming to have a large supply of these impossible squares. When you ask to see one, Rowan says: "No, it doesn't work like that. For each of our squares, I will reveal one row, and Colin one column. But you can choose which row and which column you want to know." You reply: "Do you think I was born yesterday? In each instance, Rowan can say any of his four possible answers (001, 010, 100 or 111) and Colin can choose whichever of his (000, 011, 101 or 110) do not conflict with Rowan's. For example, if I ask for the second row and the third column, and Rowan says '001', then Colin just has to choose an answer with 1 as the middle entry, either 011 or 110."

But Colin persists: "What if you prevent me from hearing not only Rowan's answer, but even the question put to Rowan? Take us far apart, and lock us in rooms that shield all forms of communication. We will still give consistent answers 100% of the time." You think to yourself: "In this case, their best strategy would be to each carry (or memorize) a predetermined list of answers to all possible questions. In a given trial, the respective answers that Rowan and Colin carry would have to correspond to squares that differ in at least one of the nine entries, because of the constraints on the rows and columns. If I conduct enough trials, choosing the questions at random, I will catch them with inconsistent answers soon enough."

So you agree to the trial as suggested; you ask questions in one room and an assistant in the other. To your consternation, Colin and Rowan give consistent answers every time. How is this possible? Are they communicating, despite all your efforts? No, they are using pairs of 'entangled' quantum particles — each pair of particles was jointly prepared in the same way, and then one kept by Rowan and one by Colin. With each trial, Rowan picks the next particle in his store, measures one of three different properties (depending on which row you ask for), and gives you one of his four possible answers based on the result of his measurement. Colin similarly processes his next particle, the one paired with Rowan's. By the 'magic' of quantum entanglement, their results are correlated precisely so as to simulate an impossible square[10].

The moral (Bell's theorem): quantum correlations falsify the hypothesis that, in any laboratory, nature carries the answer to any question which may be put there, and answers without knowing which questions are being put elsewhere. **H.W.**

and fast switching. It will be a landmark achievement in physics. But would such an experiment really close all the loopholes? The answer depends on one's attitude to causation.

The issue is whether the settings in one laboratory are uncorrelated with variables (hidden or otherwise) in the other. If they are correlated, then the experiment violates the assumptions of Bell's theorem, opening the free-choice loophole, so called because of how it can be closed: the only things correlated with free choices are their effects, so (by Einstein's principle) settings that are freely chosen late enough would be uncorrelated with the other variables, as desired.

Human choice and action are slow, so Bell experiments thus far have used random-number generators rather than free choice to change the detector settings. There is no reason for such random numbers to be correlated with anything on the other side. But if one is inclined to reject the principle of common cause (as localists are) then one must admit that correlations can occur without any reason. Thus, to be rigorous, experimenters must choose the settings freely.

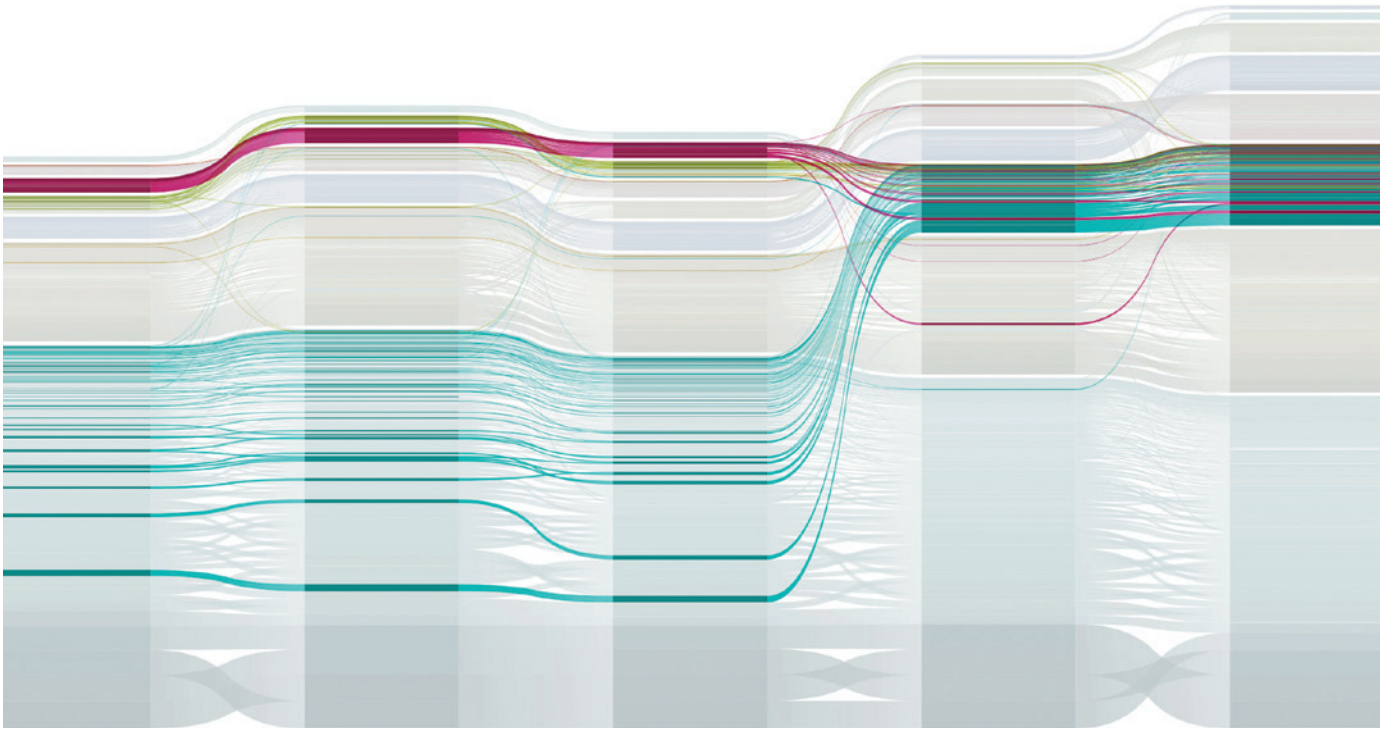
Using human free-choice while closing the separation loophole would require separating the experimenters by much more than one Earth diameter (only 40 light-milliseconds). Putting one experimenter on the Moon (1.3 lightseconds away) would also allow time for them to consciously register the results — a requirement to rule out a fourth and final loophole, the 'collapse loophole'[9]. This arises from the possibility that the set of potential results recorded by a detector does not 'collapse' to an actual individual result until observed by the experimenter, so that before the experimenter gets involved the result could be influenced, long after the photon arrives, by some bizarre (but not faster-than-light) causal influence from the distant laboratory.

Such an Earth–Moon experiment is a worthy challenge for the next 50 years. ■

Howard Wiseman *is professor of physics at the Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland, Australia.*
e-mail: h.wiseman@griffith.edu.au

1. Bell, J. S. *Physics* **1,** 195–200 (1964).
2. Aspect, A., Dalibard, J. & Roger, G. *Phys. Rev. Lett.* **49,** 1804–1807 (1982).
3. Barrett, J., Hardy, L. & Kent, A. *Phys. Rev. Lett.* **95,** 010503 (2005).
4. Maudlin, T. *Quantum Non-Locality & Relativity* (Blackwell, 2011).
5. Wiseman, H. M. Preprint at http://arxiv.org/abs/1402.0351 (2014).
6. Bell, J. S. *Epistemol. Lett.* **9,** 11–24 (1976).
7. Giustina, M. *et al. Nature* **497,** 227–230 (2013).
8. Christensen, B. G. *et al. Phys. Rev. Lett.* **111,** 130406 (2013).
9. Kent, A. *Phys. Rev. A* **72,** 012107 (2005).
10. Aravind, P. K. *Found. Phys. Lett.* **15,** 397–405 (2002).

© 2014 Macmillan Publishers Limited. All rights reserved



VISUALIZATION: MORITZ STEFANER, WITH MARTIN ROSVALL, JEVIN WEST AND CARL BERGSTROM (BERGSTROM LAB, UNIV. WASHINGTON)

Citation patterns chart, from left to right, the emergence of neuroscience as a discipline distinct from cell and molecular biology and neurology.

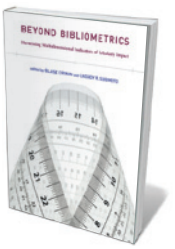
BIBLIOMETRICS

The citation game

Jonathan Adams takes the measure of the uses and misuses of scholarly impact.

Bibliometrics has a problem. In fact, the field — which tracks scholarly impact in everything from journal papers to citations, data sets and tweets — has a lot of problems. Its indicators need better interpretation, its tools are widely misused, its relevance is often questioned and its analyses lack timeliness. The expert contributors to *Beyond Bibliometrics*, edited by information scientists Blaise Cronin and Cassidy Sugimoto, rehearse the field's history and the theory behind those problems; critique both good and bad practices; and set out ideas about new indicators, including those from innovative 'altmetric' data sources. The result is a milestone for the field and a reference for practitioners. But the book misses a trick. It will not reach the non-experts — such as reviewers, hiring committees and grant panels — who use and often misuse bibliometrics to make decisions.

Bibliometrics developed during the post-1945 boom in research, expansion of publishing and development of structured electronic databases. It hit the research market 50 years ago, with Eugene Garfield's Science Citation Index, which was launched in 1964 and within years had been adopted by the US National Science Foundation. But it was in Europe, particularly in the Netherlands and Hungary in the 1980s, that citation analysis developed into research-management information. From the mid-1990s, on the back of analyses of UK research-assessment processes, bibliometrics began to enter policy-making as a way to benchmark institutions and researchers. Its commercial potential led Elsevier to create the Scopus bibliographic database, and the fundamental value of citation networks for search and discovery made Google Scholar a favourite with many researchers.



Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact
EDS BLAISE CRONIN AND CASSIDY R. SUGIMOTO
MIT Press: 2014.

Bibliometrics has had many intriguing outcomes — for instance, it has revealed the speed of China's growth and the real distribution of scientific wealth between and within nations. It helped to put research policy and management on an evidence-based footing that could challenge the assumptions and prejudices of established scientists: most countries turn out to have a smaller core of excellence than many believed, and small, agile institutions have been shown to be excellent at research, alongside older, larger centres. Rooted in the basic needs of the researcher to find the literature most pertinent to their work, this history is nicely reviewed by Nicola de Bellis, who reminds us that, for example, it is bibliometric theory — partly from Garfield and explicitly from Francis Narin and Gabriel Pinski — that led to Google's key PageRank algorithm.

Because of that historical depth of analysis, the book's handling of methods and tools goes well beyond mere numbers. For example, Jevin West and Daril Vilhena discuss research-network conceptualizations such as eigenvectors, which show how fields depend on one another. Loet Leydesdorff's comprehensive visualization of the relatedness of research disciplines reveals that variations in levels of success may depend on interdisciplinary differences in team and institutional structure. Michael Kurtz and Edwin Henneken provide a detailed analysis of the 'recommender systems' that astrophysicists use to find key literature.

One big set of problems in the field relates to careless use. The *h*-index, which captures the distribution of a scientist's citations, is innately flawed because it is not responsive to field or career stage, but its simplicity may

© 2014 Macmillan Publishers Limited. All rights reserved

have led to a plethora of derivative indices. Impact factors tell us about journals, not individual publications, but research managers use them to pick staff. Engineers publish conference proceedings, but submit only journal papers for assessment. Mathematicians publish infrequently and cite deep into the past; molecular biologists, quite the opposite. Economists and particle physicists have utterly different norms of authorship assignment. Few users of bibliometrics appreciate the implications of these nuances, yet many think they can read the numbers, or persuade others to read them their way.

Bibliometrics now underpins research policy and management and is an established part of programme evaluation for the European Commission and national agencies including the UK Research Councils. But Cronin and Sugimoto urge caution: even after decades of use, do we really understand what citation data are and what we do with them? Do those who use bibliometrics have clear criteria for how they employ and interpret them? If a tool that influences so much research is used carelessly, these arguments need to be accessible to a wide audience.

A second set of problems arises from limitations in use. Julia Lane, Mark Largent and Rebecca Rosen remind us that political paymasters and research managers want timely indicators so that they can influence what is happening rather than reflect on what has happened. But citations data pertain to work funded years before. So, in its final third, the book goes "beyond bibliometrics" to alternative metrics that may address these issues, drawn partly from social media such as Twitter and blogs (discussed by Jason Priem), and partly from the business of research, including data downloads (Kayvan Kousha, Mike Thelwall and Stefanie Haustein) and esteem indicators (such as academic genealogy, discussed by Sugimoto). Whereas citation analysis builds on decades of work, we have less than ten years of social media to analyse. So there is a tension: if thoughtful commentators dare not advance answers, instant experts rush in to satisfy policy stakeholders. If we do produce analyses built on shifting sand, and promise too much, then we could undermine what might, with care, be very informative.

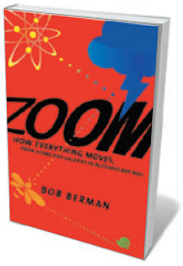
With caution and care, decisions about the future directions, funding and staffing of science will be better informed by the approaches and developments discussed in *Beyond Bibliometrics*. What we now need from this book's expert group is an accessible working narrative to guide the rest of us in our day jobs. ■

Jonathan Adams *is chief scientist at Digital Science in London and Leeds, UK, part of Macmillan Science and Education, which publishes* Nature.
e-mail: j.adams@digital-science.com

Books in brief



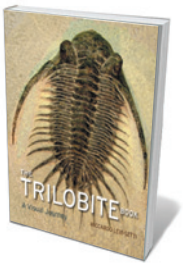
Virtual Unreality: Just Because the Internet Told You, How Do You Know It's True?
Charles Seife VIKING *(2014)*
Digital information is a "superbug of the mind" with no vaccine, notes journalist Charles Seife. Uniquely transmissible, persistent and connected, it is both an unprecedented marvel and ideally suited to misuse. Seife's analyses cut deep as he tours the 'reality' strained through the Internet's clotted filter — from the instant news churned out by online journalism to Wikipedia's shifting edits, 'sock puppetry' (fake online identities), plagiarism and viral rumours. A cogent, balanced, quietly impassioned call for Internet scepticism.



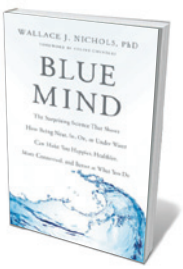
Zoom: How Everything Moves, From Atoms and Galaxies to Blizzards and Bees
Bob Berman LITTLE, BROWN *(2014)*
Science columnist Bob Berman approximates the impossible — perpetual motion — in this wit-propelled exploration of movement in nature. He bookends his tour with the "mother of all motion", the cosmos, first visiting astrophysicist Daniel Kelson (who discovered the fastest-moving known galaxy), and finally contemplating the limitless speed of stars in an infinite Universe. In between is a dazzling romp through trekking neutrinos, 60-kilometre-an-hour dragonflies, the 'kangaroo hops' of wind-blown sand grains, and more.



The Story of Pain: From Prayer to Painkillers
Joanna Bourke OXFORD UNIVERSITY PRESS *(2014)*
Pain — chronic, searing, triggered by childbirth or unrequited love — is a "definition-defying beast", avers historian Joanna Bourke. In her perceptive study covering the past three centuries, she details experiences of pain (such as "champagne bubbles and blisters" reported by a woman with a phantom limb) and responses to others' suffering. The belief that the poor and indigenous were less susceptible to pain, for instance, was still held by many into the twentieth century. We can learn to "suffer better", argues Bourke, by recognizing that pain is multifaceted and culturally embedded.



The Trilobite Book: A Visual Journey
Riccardo Levi-Setti UNIVERSITY OF CHICAGO PRESS *(2014)*
They may have died out 250 million years ago, but trilobites — a group of marine arthropods containing some 20,000 species — have a persistent grip on the human imagination. This coffee-table gem by physicist and trilobite aficionado Riccardo Levi-Setti marries the intertwined story of his global hunt for specimens and trilobites' place in prehistory with 235 superb colour photographs of select fossils. Perhaps most astounding is the array found by Arkadiy Evdokimov in Russia: their preservation is exquisite, down to the rococo flourishes of curving spines and protuberant, complex eyes.



Blue Mind: How Water Makes You Happier, More Connected and Better at What You Do
Wallace J. Nichols LITTLE, BROWN *(2014)*
The nexus of wellbeing and proximity to water — ocean shores, lakes, even fountains — holds promise as an area of multidisciplinary study, argues marine biologist Wallace Nichols. And in his popular mix of biology, neuroscience and ecosystem analysis, some findings do compel. The sight of moving water, for instance, can subtly boost daydreaming, that neural "drift" in the default-mode network that benefits brain function. **Barbara Kiser**

© 2014 Macmillan Publishers Limited. All rights reserved



NASA

Neil Armstrong trains in a simulator ahead of the *Apollo 11* Moon landing mission.

SPACE SCIENCE

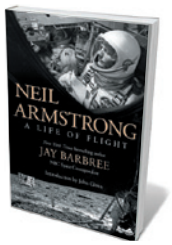
Lunar star

Roger D. Launius is perplexed by a biography of Neil Armstrong that profiles the missions, not the man.

Whatever else Jay Barbree's *Neil Armstrong: A Life of Flight* might be, it is not a biography of the first man to walk on the Moon, who died in 2012. Instead, it is an almost mission-by-mission summary of NASA's Apollo programme, observations on the nature of human spaceflight throughout the past half century and a few insightful stories. The anecdotes are detailed and effective, and sometimes illuminate larger issues. There are conversations with space celebrities, astronauts' jokes and debates over the focus of space initiatives. At the very least, this represents a unique opportunity to read the details of what some astronauts, including Armstrong, might have been thinking at times in their careers.

But what might a full biography of Armstrong have explored? First and foremost is Armstrong's multifaceted persona. He spent his life flying, becoming a pilot before he could drive. He was most proud of his naval service during the Korean War, when he flew combat missions from the *USS Essex* against targets immortalized in the James Michener novel *The Bridges at Toko-Ri* (Random House, 1953). Later, as a research pilot, he flew high-performance aircraft such as the X-15.

As an astronaut, Armstrong sought neither fame nor riches. He could have done anything he wished after he completed the *Apollo 11* Moon landing mission, yet he chose to teach aerospace engineering at the University of Cincinnati in Ohio. Perhaps no one ever really knew him — a reticence that was compelling in itself. Always gracious, Armstrong was nonplussed by the attention he received after Apollo 11; he knew that he was simply one among thousands who made the Moon landing possible. Regardless, he carried the weight of that historical mission on his back for more than 40 years. And he did much during his reclusive years after *Apollo 11*, serving on multiple national studies about spaceflight for NASA and various presidents. Although some at NASA would have preferred that he gave more public support to the agency's initiatives, Armstrong's thoughtful perspective carried weight.



Neil Armstrong: A Life of Flight
JAY BARBREE
Thomas Dunne: 2014.

I would have appreciated an explanation of the space-policy issues that Armstrong became involved in, as well as a full discussion of his role. Barbree briefly mentions a letter sent by Gene Cernan (*Apollo 17*), Jim Lovell (*Apollo 8* and *Apollo 13*) and Armstrong to US President Barack Obama in response to the space shuttle's impending retirement, but does not fully explain its content. It warned that failure to pursue an aggressive government spaceflight programme "destines our nation to become one of second- or even third-rate stature". That debate still rages. It originated in no small measure over whether to maintain the traditional approach to human spaceflight that NASA has taken for 50 years, which Armstrong apparently backed: owning the vehicles and operating them through contractors. By contrast, those from the 'new' space world want to allow private-sector firms to seize the initiative and pursue entrepreneurial approaches to human spaceflight. Traditionalists believe that the entrepreneurs will sacrifice safety; entrepreneurs point to the large, over-budget programmes of the traditionalists.

Barbree has the longest tenure of any journalist covering the space programme; his knowledge is both broad and deep. A more personal account of Armstrong would have been welcomed by all. As it is, Armstrong the man has been best captured in James R. Hansen's *First Man: The Life of Neil A. Armstrong* (Simon & Schuster, 2005). The astronaut cooperated with every aspect of that book. There will be many other fine biographies, but it is unlikely that any will be definitive, given Armstrong's complexity. ■

Roger D. Launius *is associate director of collections and curatorial affairs at the Smithsonian Institution National Air and Space Museum, Washington DC.*
e-mail: launiusr@si.edu

© 2014 Macmillan Publishers Limited. All rights reserved

Correspondence

Assess benefits and costs of shale energy

The United States and Canada are already extracting fossil fuels from shale formations by fracking, and the industry is expanding rapidly into Australia, Asia, South America and Europe. Whereas conventional energy-production sites tend to cause mainly local impacts, the wide spatial footprint of shale-energy extraction means that many more people will be affected as such sites are established.

We estimate that around 300 million people across 6 continents occupy land overlying shale-energy reservoirs (see 'Socio-economic impacts of shale energy'). This figure is based on shale gas and oil basins identified by the US Energy Information Administration (see go.nature.com/tgsmsv) and on settlement data. Large-scale industrial extraction sites are likely to have socio-economic consequences for all these people.

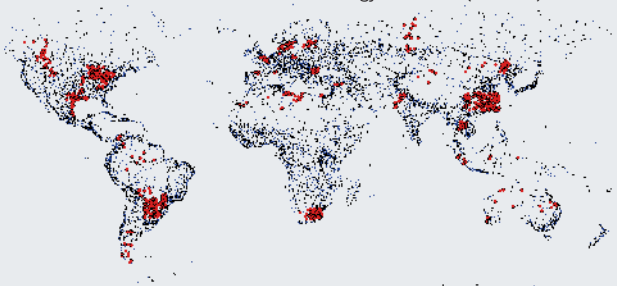
These effects could be benefits or costs, depending on how they play out. Examples include positive or negative changes to income, employment and housing costs for individuals. For communities, there could be impacts on population growth, youth migration, income distribution, the aesthetic appeal of the landscape and service provision. Extraction sites could also alter the economic productivity of host areas, for example because of competition for land use or the energy sector's high wages.

Tackling these factors fairly will be a challenge to policy-makers, who need to realize the benefits of shale-energy technology to communities. They also have to offset potential losses, including setting appropriate compensation.

Thomas G. Measham, David A. Fleming *CSIRO Ecosystem Sciences, Black Mountain Laboratories, Canberra, Australia.*
tom.measham@csiro.au

SOCIO-ECONOMIC IMPACTS OF SHALE ENERGY

Of all humans in populated areas (dark blue dots) across the globe, some 300 million live above known shale-energy reservoirs (red dots).



DATA SOURCES: NATURALEARTHDATA.COM/EIA

Maize is not a clue to Puerto Rican origins

You discuss the implications of the unpublished discovery of DNA from maize (corn) in ancient faeces from two pre-Columbian cultures on Vieques Island, Puerto Rico (*Nature* http://doi.org/s8g; 2014). We question the inference from this discovery that one culture, the Huecoid people, originated from the Bolivian Andes.

That inference is presumably based on the assumption that maize and chicha (a fermented maize beverage) were produced in Bolivia. However, both items were widely used throughout pre-Columbian America during Huecoid times in 400 BC to AD 600 (see, for example, R. M. Bonzani and A. Oyuela-Caycedo in *Histories of Maize* 343–356; Academic Press, 2006). Maize was also already present among pre-Arawak Antillean groups by 2950 BC and, contrary to your speculations, was not introduced into Puerto Rico by the Huecoid people (see J. R. Pagán-Jiménez in *The Oxford Handbook of Caribbean Archaeology* 391–406; Oxford Univ. Press, 2013).

The archaeological evidence of early Caribbean societies therefore indicates that Puerto Rican natives may have had earlier and more wide-ranging origins than your report implies (see also R. Rodríguez-Ramos *Rethinking Puerto Rican Precolonial History;* Univ. Alabama Press, 2010).

Jaime R. Pagán-Jiménez *Leiden University, the Netherlands.*
Reniel Rodríguez-Ramos *University of Puerto Rico, Utuado, Puerto Rico.*
José R. Oliver *University College London, UK.*
j.r.pagan.jimenez@umail.leidenuniv.nl

Greek science must reform to survive

Greece's progress in science was remarkable in the decade before the country's severe economic downturn at the end of 2008. Although subsequent research output has been sustained against the odds, prompt and coordinated action is now needed to avert a crisis.

During 1999–2009, the proportion of papers published worldwide that included at least one Greek author rose from 0.43% to 0.76%, and Greece moved up from 31st to 24th in the global ranking of annual publications (data from Scopus, derived using SCImago; www.scimagojr.com). At the peak of its research output in 2009, Greece had a greater scientific output per capita than France, Italy or Japan — despite the Greek government's record of under-investment in research.

Because Greece's research budget has since fallen behind that of the rest of Europe, this peak has become a plateau — unlike in almost all other European Union countries, where publication output is rising. Hiring of staff has ceased and researchers' salary cuts are killing competitiveness and leading to a mass exodus to foreign labs.

To avoid a descent back to last century's rankings, it is vital that Greece uses its resources creatively and efficiently by instigating essential structural reforms, including a national science plan, evaluation of funding allocation, and university restructuring. It should also encourage stronger cooperation among national research centres and more international collaboration.

Nikolaos Vrachnis, Nikolaos Vlachadis *University of Athens Medical School, Athens, Greece.*
Dionisios Vrachnis *University of Patras Medical School, Greece.*
nvrachnis@hotmail.com

Update German view of genetic testing

There has been scant public discussion about genetic screening of embryos in Germany, presumably because of a residual suspicion of genetic diagnostics after the sinister history of Nazi eugenics. To rectify this and to remove ideological preconceptions, the Leopoldina, Germany's national academy of sciences, set up an interdisciplinary panel of experts last year to discuss the ethical implications of genetic diagnostics (see go.nature.com/gzmhpe; in German).

The panel agreed that adequate genetic counselling is paramount, that personal freedom and human dignity must be preserved, and that a third party should not be allowed to decide the fate of an embryo that tests positive for a severe genetic disease.

The panel's main concerns were to find the best ways to safeguard the voluntary decisions of couples to undergo and act on genetic testing of their unborn child, and to protect the rights of tested and untested children born with a genetic disease.

Peter Propping, Heinz Schott *University of Bonn, Germany.*
propping@uni-bonn.de

© 2014 Macmillan Publishers Limited. All rights reserved

Gerald Edelman

(1929–2014)

Biologist who won Nobel for solving antibody structure.

Gerald Edelman was a man of diverse talents and broad knowledge. A pre-eminent biologist, he was also an accomplished writer and musician, and read widely in philosophy. This brought an elegant style to his science, and gave his more speculative intellectual works a characteristic literary format.

Edelman died on 17 May in La Jolla, California. Born in Queens, New York, in 1929, he studied medicine. He then turned to immunology and spent most of his research career at Rockefeller University in New York. His focus on the structure of antibodies in the 1960s led to the 1972 Nobel Prize in Physiology or Medicine, which he shared with Rodney Porter. The award recognized Edelman's work on the descriptions of the heavy (H) and light (L) polypeptide chains, and Porter's on the distinct binding domain (antigen-binding or Fab fragment) of antibodies (also called immunoglobulins). These were the first dominoes to fall in the elucidation of the structural variability that lies at the heart of antibody-based recognition of infectious agents such as bacteria and viruses.

With colleague Joseph Gally, Edelman discovered that L-chain polypeptides were secreted in the urine of a person with a myeloma tumour. Being homogeneous (that is, each chain contained a single amino acid sequence), these light chains were amenable to structural characterization, unlike the mixture of antibodies in blood.

After the prize, Edelman's group expanded its interests into several areas, including carbohydrate-binding proteins, cell-to-cell adhesion mediated by proteins on the surface membranes of cells, and tissue morphogenesis. But his personal focus was turning towards the nervous system, leading in 1981 to his founding and direction of the Neurosciences Institute in New York, which he moved with him in 1993 to San Diego, California, where it became part of the Scripps Research Institute (from which it became independent in 2012).

Edelman was fascinated throughout his career by the enormous explanatory potential of selective-recognition systems. These had already been used to attack and elucidate two of the most challenging problems in biology: the evolution of species and the adaptive immune response. Darwinian evolution of species depends on fitness-based selection from a repertoire of variable genetic traits; the immune system recognizes foreign substances by selecting from a repertoire of antigen-recognizing cells.

Over about the same period as the antibody studies, two fundamental properties of neurons and their synaptic connections were discovered by different teams: that cortical neurons are organized into discrete groups of cells, and that synapses strengthen through use. Edelman reasoned that the formidable recognition and processing capabilities of a complex nervous system depend on selection operating on cell groups that differ in their connectivity patterns. Incoming sensory information would elicit a response from distinct cell groups; it could be modified, he posited, by repetitive recognition (termed re-entry) that provided opportunities for strengthening, abstraction and association. Thus, he argued, the simplicity of repeated selection from an evolving repertoire could account for the astonishing information-processing power of sophisticated neural systems.



CLAUDIO VITALE

Edelman presented this idea in several books, including *Neural Darwinism* (Basic, 1987), *Bright Air, Brilliant Fire* (Basic, 1993) and *A Universe of Consciousness* (Basic, 2001), which was written with colleague Giulio Tononi. The descriptions and arguments are often dense and elaborate, viewed by some readers as possessing an admirable literary quality, and by others as redundant and opaque.

Given his multidimensional background and relentless intellect, it is not surprising that Edelman was a complicated and controversial individual, in his research and his relationships. He could range from charming and inspiring to arrogant and abrupt. Although colleagues often found him abrasive, in his group Edelman could be an effective and generous leader. He sustained collaborations with several prominent scientists, including Gally, Bruce Cunningham — whose biochemical skills were essential in the formidable task of antibody analysis in the 1960s and 70s — and George Reeke in the early modelling of selective brain function from the 1980s onwards.

Edelman was also an exceptional mentor. He had a gift for formulating important hypotheses. He instilled a pragmatic attitude towards experimental design: his coinage 'the refrigerator principle' expressed that ideas sometimes come from the ready availability of reagents. He warned of the dangers of distraction — 'don't look sideways' — and stressed the value of interpreting results in a broader context.

Training with him was intense. He once told a senior student: "It's your year, son, to be in the nitric acid sitz bath." Thesis presentations in Rockefeller's legendary Caspary Hall were preceded by weeks of his direct scrutiny — experiences of enduring value to building a research career, and becoming a mentor for the next generation.

With Edelman's passing, there remains the task of testing his theory of selection-based brain function. As with Darwin's species evolution and the adaptive immune response, the theory was proposed before anyone understood the underlying cellular and molecular mechanisms, and before there were adequate analytical tools. There have been some remarkable advances in the past five years in the use of genetics and molecular reporters of synaptic activity that allow real-time evaluation of substantial groups of cells in functioning brain tissue. Thus one can begin to imagine a line of investigation that could address the challenge that his provocative work presents to neuroscience. ■

Urs Rutishauser *is at the Sloan Kettering Institute at the Memorial Sloan Kettering Cancer Center in New York. He was a student and colleague in Gerald Edelman's laboratory at Rockefeller University in New York from 1967 to 1983.*
e-mail: u-rutishauser@mskcc.org

© 2014 Macmillan Publishers Limited. All rights reserved

FORUM: Nuclear-weapons dismantlement

Identifying a hidden warhead

A means of verifying that nuclear warheads to be dismantled are genuine items has been proposed that potentially reveals no information to an inspector about the design of the weapons. Two experts explain the ins and outs of the method and its implications for arms-control policy. SEE ARTICLE P.497

THE PAPER IN BRIEF

- To prevent the proliferation of nuclear weapons, the protocol used by an inspector to verify the authenticity of a weapon must not reveal sensitive information about the weapon's design.
- This requirement presents a major challenge, because currently proposed inspection procedures may allow tampering with instrumentation and data.
- Glaser *et al.*[1] (page 497) have developed a protocol based on neutron transmission in which sensitive information is not actually measured.
- The procedure potentially meets the restriction on information disclosure and could allow inspectors to verify weapon authenticity with high confidence.

Nuclear blind man's buff

JOHN FINNEY

How do you know that an item concealed in a box is what you are told it is if you are not allowed to measure the properties of the item through the box walls? This is the problem faced by inspectors in satisfying themselves that the item is a nuclear warhead. Instruments based on γ-ray and neutron spectroscopy could tell us that the box contains a certain amount ('*x*' kilograms) of plutonium-239. But because such measurements would reveal warhead-design information that could assist weapon proliferation, their full use contravenes the Treaty on the Non-Proliferation of Nuclear Weapons.

The currently accepted solution is to pass the data from the instrument's detector through a 'redactive' information barrier so that the inspector receives only non-proliferative information (Fig. 1). Ideally, this would be a yes or no answer to questions such as, 'does the mass of plutonium exceed *x* kg', or 'does its isotopic composition exceed *y*% of ^{239}Pu'? Work to try to achieve such outcomes has ranged from software or hardware data manipulation to systems based on neural networks, genetic algorithms and pattern recognition[2–4]. But the (some would say) Byzantine weapon-dismantling procedures required to maintain security provide opportunities for tampering with instrumentation and data by both host and inspector.

Glaser and colleagues' study gets over the problem of the inspector seeing classified information by just not measuring it. This may seem strange, but such a 'zero-knowledge' approach has a good pedigree in cryptography[5]. In essence, the proposed protocol, which is based on neutron transmission through the contained warhead, gives a null result if the item is what the host says that it is.

The authors' neat explanation of the principle illustrates how a simple technique — counting marbles — applied with logic can give a high degree of confidence that two cups of marbles contain the same (but unknown) number of marbles (see Fig. 1 of the paper[1]). In the neutron case, counting neutrons transmitted through the warhead will produce a neutron radiograph (call this T); this is similar to a medical X-ray, only using neutrons rather than X-rays. To avoid seeing this radiograph, and hence obtaining proliferative information, the inspector uses detectors that are preloaded by the host with the negative of the contained warhead's radiograph ($-T$). So, if the item is genuine the inspector will see a null signal: $-T+T=0$. He or she can then confirm that the item is as declared, but is no wiser about its detailed properties.

There are problems to be addressed before such a protocol can be implemented, and Glaser and colleagues discuss several of these in their article. But as long as such problems can be overcome, this brilliant idea could solve one of the really messy issues that we will have to deal with when we finally agree to rid ourselves of these weapons.

John Finney *is in the Department of Physics and Astronomy and the London Centre for Nanotechnology, University College London, London WC1E 6BT, UK.*
e-mail: john.finney@ucl.ac.uk

Waiting for politics to catch up

JAMES M. ACTON

The problem seems as imponderable as Bilbo's riddle to Gollum in *The Hobbit*: "What have I got in my pocket?" A state has a concealed object that it claims is a nuclear weapon. It wishes to provide proof, but the object's design is largely secret and inspectors are forbidden from making measurements that might reveal classified information. Glaser and colleagues' ingenious solution is an unclassified measurement that provides statistical

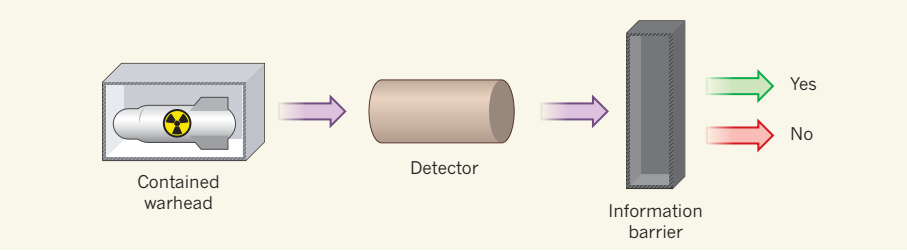


Figure 1 | Conventional warhead-verification scheme. The usual approach to verifying that a nuclear warhead concealed in a container is what an inspector is told it is involves an information barrier. This barrier accepts high-quality information from a detector, processes it in such a way that classified data about the concealed item do not pass through to the inspector, and passes on a simple yes or no answer to an agreed fact about the item. Glaser *et al.*[1] propose a new protocol to nuclear-warhead verification that does not involve such a barrier but potentially reveals no classified information.

© 2014 Macmillan Publishers Limited. All rights reserved

confidence that the object is identical to an agreed 'template'. Unfortunately, as is often the case, the technology of arms control is some years ahead of its politics.

The most recent arms-control agreement between Russia and the United States, the New Strategic Arms Reduction Treaty (New START), limits deployed strategic warheads — those physically attached to long-range delivery systems. For verification purposes, it permits inspectors to count objects inside the nose cone of a ballistic missile. However, inspectors do not check that these objects really are warheads. After all, if Russia wants to put a lump of rock on a missile and pretend that it is a warhead, the United States is not disadvantaged.

Once warheads are removed from missiles, however, they leave New START's accountancy system. They can be — and frequently are — stored rather than destroyed. As a result, arms controllers have proposed that the United States and Russia negotiate a new agreement covering warheads in storage. The administration of US President Barack Obama has endorsed this idea. However, even before Russia's annexation of Crimea in March this year, there was no short-term path to clearing the obstacles, including a particularly vexatious disagreement over ballistic-missile defence, that lie in the way of a new treaty[6].

In the future, changing political circumstances may provide an opening for a more ambitious treaty that limits all warheads and requires the destruction of any excess. In this case, unlike under New START, a state might have an incentive to cheat by trying to pass off a 'fake' as an authentic warhead slated for dismantlement, allowing the real thing to be diverted to a clandestine stash. Glaser and colleagues' approach provides a promising way for inspectors to ensure that they are not being hoodwinked.

While waiting for political change, government scientists could subject this protocol to an independent analysis to confirm that inspectors would not learn classified information, and non-governmental experts could compare this protocol with other, more conventional approaches[7]. After all, we should probably work out how to verify a comprehensive treaty before negotiating one. ■

James M. Acton *is at the Carnegie Endowment for International Peace, Washington DC 20036-2103, USA.*
e-mail: jacton@ceip.org

1. Glaser, A., Barak, B. & Goldston, R. J. *Nature* **510,** 497–502 (2014).
2. *Discovery* No. 23, 36–47 (AWE, 2012).
3. Anderson, B. *et al. Verification of Nuclear Weapon Dismantlement: Peer Review of the UK MoD Programme* (British Pugwash, 2012).
4. Aitkenhead, M. J., Owen, M. & Chambers, D. M. *J. Anal. Atom. Spectrom.* **27,** 432–439 (2012).
5. Chazelle, B. *Nature* **446,** 992–993 (2007).
6. Acton, J. M. *Wash. Q.* **35**(2), 37–53 (2012).
7. *Monitoring Nuclear Weapons and Nuclear-Explosive Materials: An Assessment of Methods and Capabilities* 97–108 (Natl Acad. Press, 2005).

ANTIBIOTIC RESISTANCE

To the rescue of old drugs

A naturally occurring fungal compound has been found to restore the susceptibility of bacteria to a class of antibiotic that is currently considered to be our last defence against serious infections. SEE ARTICLE P.503

DJALAL MEZIANE-CHERIF & PATRICE COURVALIN

The emergence and spread of antibiotic resistance in bacteria has potentially drastic consequences for human health around the world. There is a pressing need for new antibiotics, or for further strategies to overcome antibiotic resistance — not only to ensure that we can continue to treat infectious diseases, but also to safeguard the practice of medical procedures such as organ transplants, chemotherapy and surgery, which are accompanied by an increased susceptibility to infections. One approach to combating resistance is to use combined treatments consisting of an antibiotic and a compound that potentiates the antibiotic's activity by targeting bacterial resistance mechanisms[1]. On page 503 of this issue, King *et al.*[2] report that aspergillomarasmine A, a naturally occurring fungal product, is a rapid and potent inhibitor of enzymes that confer antibiotic resistance on enterobacteria.

There has been growing alarm over the spread of bacterial resistance to carbapenems[3] (antibiotics containing a β-lactam ring structure that is essential for their function). Carbapenems are used to treat serious infections due to enterobacteria such as *Escherichia coli* and *Klebsiella pneumoniae*, which cause a variety of conditions, including pneumonia, septicaemia and meningitis.

Enterobacteria become resistant to carbapenems by producing one of two types of enzyme — serine β-lactamases or metallo-β-lactamases (MBLs). Serine β-lactamases contain a serine amino-acid residue that inactivates the antibiotic by cleaving the drug's β-lactam ring. MBLs contain a zinc ion that activates a nucleophilic water molecule, leading to cleavage of the ring by hydrolysis (Fig. 1a). Several inhibitors of serine β-lactamases are already used in combination with penicillins to overcome resistance[4] (one commonly used example is Augmentin, a combination of clavulanic acid and amoxicillin).

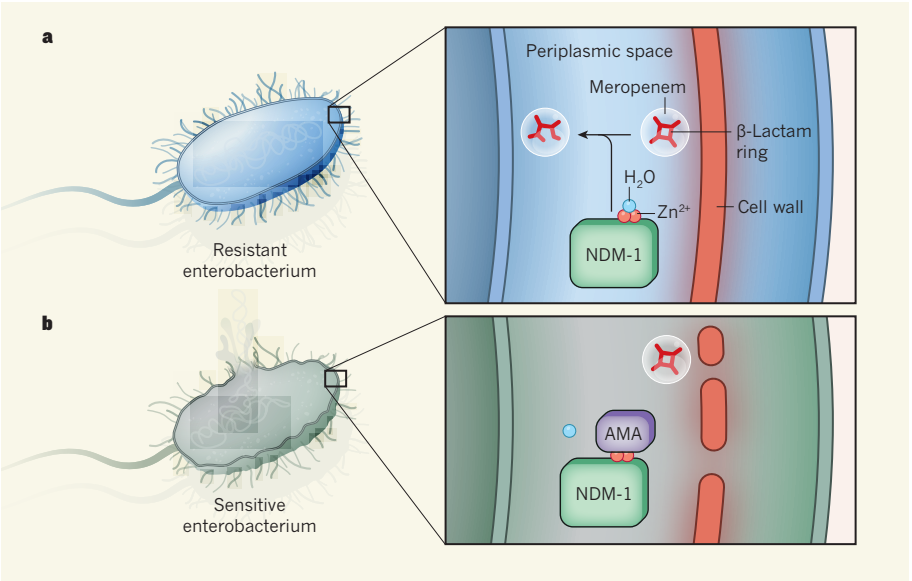


Figure 1 | Overcoming antibiotic resistance. **a**, Production of the enzyme NDM-1 by enterobacteria confers resistance to antibiotics of the carbapenem family. One such antibiotic is meropenem, which acts in the periplasmic space, targeting the cell wall. NDM-1 inactivates meropenem through zinc ions, which activate a nucleophilic water molecule, leading to cleavage of the antibiotic's β-lactam ring. **b**, King *et al.*[2] report that a naturally occurring fungal compound, aspergillomarasmine A (AMA), can interact with the zinc ions of NDM-1, preventing inactivation of meropenem, which then kills the enterobacteria by breaking down their cell wall.

© 2014 Macmillan Publishers Limited. All rights reserved

However, despite several efforts[5], no effective inhibitors of MBLs are available.

Most members of our current antibiotic arsenal originate from screens of naturally occurring chemicals produced by soil microorganisms[6]. King and colleagues reasoned that replication of this approach might lead to the discovery of small molecules that could resensitize bacteria to drugs against which they have developed resistance. The authors performed a screen of naturally occurring microbial extracts to find compounds that could inhibit an MBL called New Delhi metallo-β-lactamase 1 (NDM-1).

Their screen identified one compound, aspergillomarasmine A (found in an extract from the fungus *Aspergillus versicolor*), that could restore the efficiency of meropenem, a carbapenem, against *E. coli* strains producing NDM-1. Aspergillomarasmine A has previously been shown to inhibit angiotensin-converting enzyme (ACE), which is produced in the body and causes blood-vessel constriction, increasing blood pressure. ACE is a zinc-dependent metalloproteinase and shares some functional similarities with MBLs, which suggests that aspergillomarasmine A might inhibit NDM-1 by interacting with zinc ions. Indeed, the authors found that aspergillomarasmine A removes zinc ions from NDM-1 in the same fashion as other known metal-binding molecules that interact with MBLs[7] (Fig. 1b).

King and colleagues tested the ability of their compound to combat resistance in 229 MBL-expressing strains of bacteria that had been isolated from patients over 10 years, from various parts of the globe. When used in combination with meropenem, they found that aspergillomarasmine A restored antibiotic susceptibility in 88% of NDM-1-producing strains of enterobacteria. Remarkably, in mice infected with a lethal strain of *K. pneumoniae* that produces NDM-1, a single-dose treatment of aspergillomarasmine A and meropenem led to a survival rate of more than 95% after 5 days of infection.

The authors showed that aspergillomarasmine A is selective — it has potent activity against NDM-1 and another MBL, Verona integron-encoded metallo-β-lactamase 2 (VIM-2), but much weaker activity against ACE. Unfortunately, aspergillomarasmine A is also ineffective against other anti-carbapenem enzymes, which, if this treatment were rolled out, might become more common, leading to resistance again. Furthermore, treatments involving compounds that aim to remove zinc ions from MBLs might inhibit the body's own metalloenzymes, causing serious side effects. However, a previous study[8] demonstrated that mice treated with aspergillomarasmine A had few side effects, and on average showed no change in arterial blood pressure. Although this is encouraging, the effects on other mammalian enzymes should be tested before drugs based on this premise are developed.

King and co-workers' study re-emphasizes the fact that drug–inhibitor combinations can be used to address the problem of increasing antibiotic resistance, through extending the lifespan of existing antibiotics. Theoretically, this approach could be applied to any resistance mechanism that affects the treatment of disease. However, many pathogens harbour more than one mechanism that confers resistance to a given drug class[9]. For example, efflux pumps in the cell membrane remove toxic chemicals (including most classes of antibiotic) from the cell, and a single inhibitor cannot address this problem as well as other resistance mechanisms, such as modification of the antibiotic's target or enzymatic degradation of the drug. Such multi-pronged resistance means that the future of antimicrobial chemotherapy will have to rely on combinations of drugs with different targets, as is the case for therapies for cancer and viral infections.

Designing new antibacterial drugs is not an easy task, requiring the development of synthetic chemicals combined with exploration of naturally occurring compounds. But the reservoir of natural products with the potential to act as antibacterial drugs has not yet been exhausted. In contrast to general thinking by drug companies, screening for such products may well still have a bright future. ■

Djalal Meziane-Cherif *and* **Patrice Courvalin** *are in the Unité des Agents Antibactériens, Institut Pasteur, Paris 75724, France.*
e-mails: djalal.meziane-cherif@pasteur.fr; patrice.courvalin@pasteur.fr

1. Kalan, L. & Wright, G. D. *Expert Rev. Mol. Med.* **13,** e5 (2011).
2. King, A. M. *et al. Nature* **510,** 503–506 (2014).
3. Woodford, N., Wareham, D. W., Guerra, B. & Teale, C. *J. Antimicrob. Chemother.* **69,** 287–291 (2014).
4. Drawz, S. M. & Bonomo, R. A. *Clin. Microbiol. Rev.* **23,** 160–201 (2010).
5. Fast, W. & Sutton, L. D. *Biochim. Biophys. Acta* **1834,** 1648–1659 (2013).
6. Lewis, K. *Nature Rev. Drug Discov.* **12,** 371–387 (2013).
7. Siemann, S. *et al. Biochim. Biophys. Acta* **1571,** 190–200 (2002).
8. Matsuura, A. *et al. Jpn. J. Pharmacol.* **63,** 187–193 (1993).
9. Courvalin, P., Leclercq, R. & Rice, L. B. (eds) *Antibiogram* (ASM Press, 2010).

FUNDAMENTAL CONSTANTS

A cool way to measure big *G*

Published results of the gravitational constant, a measure of the strength of gravity, have failed to converge. An approach that uses cold atoms provides a new data point in the quest to determine this fundamental constant. SEE LETTER P.518

STEPHAN SCHLAMMINGER

In our daily lives, we can see the effect of the gravitational force between Earth and an object, say an apple. However, the gravitational attraction between two apples cannot be observed without using a sensitive apparatus such as a torsion balance — it is just too small. On page 518 of this issue, Rosi *et al.*[1] describe an amazing measurement: the gravitational force between a rubidium atom and a 516-kilogram mass, with a relative uncertainty of just 0.015%. Their experiment was aimed at a precise determination of the gravitational constant, which describes the strength of the gravitational pull that bodies exert on each other, and was based on the technique of atom interferometry — a method that takes advantage of the wave nature of cold atoms to precisely measure gravitational acceleration.

In the atom interferometer described by Rosi and colleagues, a cloud of rubidium atoms at a temperature close to absolute zero is repeatedly tossed up vertically. To understand how this cloud in free fall probes gravity, quantum mechanics is needed. For simplicity, consider that the atoms in the cloud can be in two different atomic states, A and B. At the beginning, all atoms are in state A. By exposing an atom to an appropriately shaped light pulse, the atom can transition from A to B with a certain probability, let's say 50%. While the atom is not being observed, it is simultaneously in both states (50% in A and 50% in B), a concept known as superposition. In addition to inducing the transition from A to B, the light pulse transfers vertical momentum such that the B state has a larger vertical velocity than the A state.

The relative fraction of the two different states in this superposition varies with time, and its rate of change depends on the difference of the products of the momentum and the travelled vertical distance for each state. Owing to its larger momentum, state B travels higher than state A in the presence of the local gravitational acceleration, g, caused mostly by Earth and any masses in the vicinity of the cloud. Hence, the rate of change is a function of g. After the atomic cloud descends, close to the

© 2014 Macmillan Publishers Limited. All rights reserved

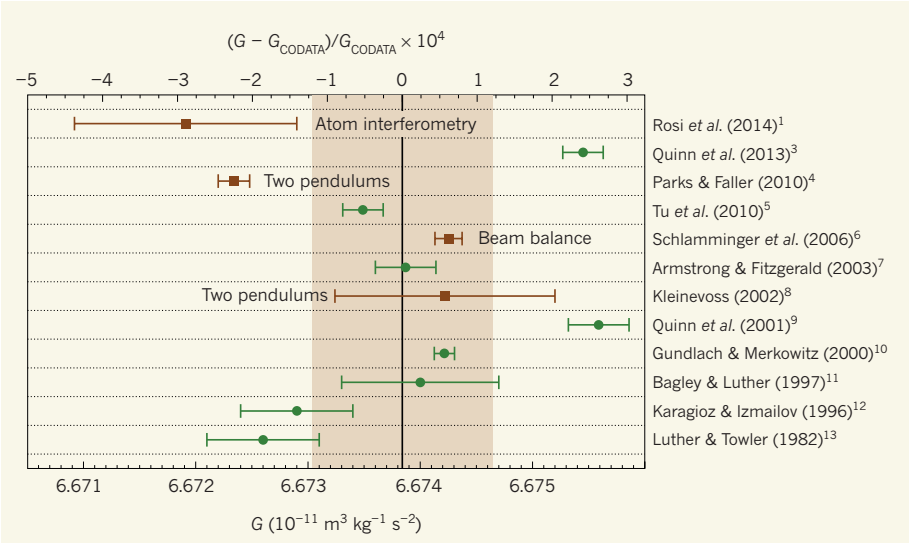


Figure 1 | The big picture for big *G*. Published results of measurements of the gravitational constant, *G*, over the past 32 years. The solid circles denote measurements that employed torsion balances. The three lower solid squares show results that were obtained using a beam balance or two pendulums. The upper solid square is the result obtained by Rosi and colleagues[1] using the technique of atom interferometry. The shaded area denotes the one-standard-deviation confidence interval of the value from the 2010 CODATA compilation of physical constants[18].

launch point, the ratio of the number of atoms in state A to state B is measured, from which g can be calculated[2].

To measure the gravitational constant G, an external mass, referred to as a field mass, is required. To understand the principle of the experiment we make two simplifications: we assume the field mass is a point with mass M, and that the atom interferometer measures g at one fixed point. In reality, the atom interferometer measured along a trajectory and cylindrical field masses were used. In this case, the idea is the same, although the maths is more complicated. The point mass is first located a distance z above the interferometer and the acceleration $g_{above} = -g + G(M/z^2)$ is measured. Next, the point mass is moved to a distance z below and $g_{below} = -g - G(M/z^2)$ is obtained. As long as g remains the same between the two measurements, G can be obtained from the difference between the measurements, $G = (g_{above} - g_{below})\, z^2/(2M)$. Unfortunately, g changes with time owing to tidal accelerations produced by the Sun and the Moon, air-pressure variations, and the movement of masses in the vicinity of the experiment.

To solve this problem, Rosi and co-workers measured g_{above} and g_{below} simultaneously by stacking two atom interferometers on top of each other. Two field masses were used and were at first in between the interferometers. The measured difference between g_{above} and g_{below} (the signal) is mostly independent of the temporal variation of g, but is dependent on its spatial variation, because the measurements were taken at different locations. The field masses were then moved such that one was above the upper interferometer and the other was below the lower interferometer, and the measurement was repeated. The difference between the signals in the latter field-mass configuration and the former one is independent of the spatial variation of g, and G was obtained by averaging about 100 such signal differences. The result is $G = (6.67191 \pm 0.00099) \times 10^{-11}$ cubic metres per kilogram per square second. The relative uncertainty of the measurement is 0.015%.

The experiment is exciting because it uses modern tools to solve an old problem. Using atoms to sense gravity instead of conventional mechanical devices, such as torsion balances, has several advantages. For example, the atom does not require a physical connection to the laboratory and is hence not biased by stray forces that such a connection would introduce to the measurement.

Naively, one would think that torsion balances are much better tools to precisely determine G than other devices. The torsion balances are much simpler in design and measure in a direction perpendicular to g, avoiding systematic effects caused by temporal variations of g. However, measurements performed over the past two centuries, mostly using torsion balances, have failed to converge on a trustworthy value of G. Figure 1 shows the results obtained in the past three decades. Out of the 11 results[3–13] shown, only three were measured with devices other than a torsion balance. One measurement was performed with a beam balance, a device that is typically used to measure mass, and two with pairs of pendulums. The relative difference between the largest and the smallest number is 0.055% — or about 40 times the size of the error bars of the experiment with the smallest uncertainty.

The various measurements of G seem not to converge on a value; it seems that the convergence gets worse with each additional data point. This is especially disconcerting because it is thought that G is a fundamental constant of nature. Although we cannot rule out for certain that the spread of the obtained values is caused by so far undiscovered properties of gravitation, this hypothesis seems unlikely because most plausible modifications to our theory of gravitation are excluded by other experimental tests. Adding more data points from isolated experiments has not been the best strategy to improve the situation. Instead, forming an international consortium[14] to coordinate these demanding experiments has been suggested.

Under the auspices of such a consortium, one or more apparatus can be sent to different institutions for measuring G. The different results and uncertainties could be compared. Such a procedure would provide insight into underestimation of uncertainties, the propensity to overlook bias in the experiment, and 'intellectual phase locking'[15], which is the tendency of an experimenter to stop looking for systematic effects once the measurement agrees with previously published results. By enhancing our understanding of these three human sources of error, which could be responsible for the spread shown in Figure 1, a more credible value of G can be obtained.

Rosi and colleagues' experiment provides an important data point in our quest to measure G (ref. 16). The experiment is vastly different from all other measurements, and the size of the achieved uncertainty, although still somewhat large, is approaching those obtained using torsion balances. Over the past 6 years, this team has reduced the uncertainty of their experiment by a factor of 10 compared with a preliminary result published in 2008 (ref. 17). Stay tuned, as they continue to push this technique to smaller uncertainties. ■

Stephan Schlamminger *is at the National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA.*
e-mail: stephan.schlamminger@nist.gov

1. Rosi, G., Sorrentino, F., Cacciapuoti, L., Prevedelli, M. & Tino, G. M. *Nature* **510,** 518–521 (2014).
2. Kasevich, M. & Chu, S. *Phys. Rev. Lett.* **67,** 181–184 (1991).
3. Quinn, T., Parks, H., Speake, C. & Davis, R. *Phys. Rev. Lett.* **111,** 101102 (2013).
4. Parks, H. V. & Faller, J. E. *Phys. Rev. Lett.* **105,** 110801 (2010).
5. Tu, L.-C. *et al. Phys. Rev. D* **82,** 022001 (2010).
6. Schlamminger, S. *et al. Phys. Rev. D* **74,** 082001 (2006).
7. Armstrong, T. R. & Fitzgerald, M. P. *Phys. Rev. Lett.* **91,** 201101 (2003).
8. Kleinevoss, U. PhD thesis, Univ. Wuppertal (2002).
9. Quinn, T. J., Speake, C. C., Richman, S. J., Davis, R. S. & Picard, A. *Phys. Rev. Lett.* **87,** 111101 (2001).
10. Gundlach, J. H. & Merkowitz, S. M. *Phys. Rev. Lett.* **85,** 2869–2872 (2000).
11. Bagley, C. H. & Luther, G. G. *Phys. Rev. Lett.* **78,** 3047–3050 (1997).

© 2014 Macmillan Publishers Limited. All rights reserved

12.Karagioz, O. V. & Izmailov, V. P. *Izmer. Tekh.* **10,** 3–9 (1996).
13.Luther, G. G. & Towler, W. R. *Phys. Rev. Lett.* **48,** 121–123 (1982).
14.http://pml.nist.gov/bigg
15.Branscomb, L. M. *Am. Sci.* **73,** 421–423 (1985).
16.Quinn, T. *Nature* **505,** 455 (2014).
17.Lamporesi, G. *et al. Phys. Rev. Lett.* **100,** 050801 (2008).
18.Mohr, P. J., Taylor, B. N. & Newell, D. B. *Rev. Mod. Phys.* **84,** 1527–1605 (2012).

This article was published online on 18 June 2014.

STRUCTURAL BIOLOGY

Wobble puts RNA on target

Enzymes that attach amino acids to transfer RNAs during protein synthesis must recognize both substrates specifically. Crystal structures reveal a mechanism that explains the RNA specificity for one such system. SEE ARTICLE P.507

OSCAR VARGAS-RODRIGUEZ & KARIN MUSIER-FORSYTH

How genetic information encoded in nucleic acids is translated into amino-acid sequences in proteins has fascinated researchers for decades. Faithful translation involves specific pairing of triplets of nucleotides, called codons, in messenger RNA with complementary anticodons in transfer RNA. High accuracy also requires the attachment (aminoacylation) of each type of amino acid to a specific type of tRNA. Each variant of the aminoacylation reaction is catalysed by a particular aminoacyl-tRNA synthetase (ARS) enzyme, which must distinguish its own amino-acid substrate and the matching tRNA from the large pool of similar substrates. In this issue, Naganuma *et al.*[1] (page 507) report crystal structures that cast light on the remarkable mechanism by which information encoded in part of a tRNA is interpreted by an ARS, resulting in highly specific aminoacylation.

One of the simplest examples of tRNA recognition is that of alanyl-tRNA synthetase (AlaRS), the enzyme that attaches the amino acid alanine to its tRNA ($tRNA^{Ala}$). Across all domains of life, this recognition process relies primarily on a single 'wobble' base pair (designated G3·U70, where G and U represent the bases guanine and uracil, respectively) in a region of $tRNA^{Ala}$ called the acceptor stem[2,3]. The AlaRS·tRNA system has long been a model for studies of protein–RNA interactions and has been extensively investigated. Much was learned from these studies and from defining the three-dimensional structures of most of the other ARS·tRNA complexes, but a crystal structure of the AlaRS·$tRNA^{Ala}$ complex has been lacking. Naganuma and colleagues now report the crystal structure of AlaRS from the microorganism *Archaeoglobus fulgidus* in complex with $tRNA^{Ala}$/GU (the wild-type tRNA that contains the G3·U70 wobble pair), and that of a defective complex involving a mutant $tRNA^{Ala}$ that contains a standard Watson–Crick pair in place of the wobble pair ($tRNA^{Ala}$/AU).

Nucleotides that mark a tRNA for specific aminoacylation are known as tRNA identity elements, and the identity sets of tRNA-synthetase systems in bacteria are well understood[4,5]. In addition to the acceptor-stem domain, which is close to the site of amino-acid attachment, anticodons form part of the identity set for most tRNAs. The importance of the G3·U70 pair in AlaRS recognition, and the fact that the enzyme does not recognize the anticodon, helped to explain the early finding[6] that a $tRNA^{Ala}$ fragment derived from the acceptor stem is a substrate for aminoacylation by AlaRS, and led to speculation that a 'second genetic code' might determine the amino-acid specificity of tRNA molecules[7].

It was subsequently observed[8] that replacement of the wobble base pair with Watson–Crick base pairs greatly reduced the rate of catalysis (k_{cat}) of the aminoacylation of $tRNA^{Ala}$, rather than the binding affinity of AlaRS for $tRNA^{Ala}$. Specific chemical groups in and around the G3·U70 base pair[9,10] and distortion of the acceptor-stem helix of $tRNA^{Ala}$ (refs 11, 12) were also shown to have a role in recognition. These results left open the question of how discrimination by AlaRS was achieved largely through its effects on k_{cat}.

Naganuma and co-workers' structures reveal that AlaRS forms a homodimer (in which two identical subunits of the enzyme associate), and that each homodimer binds a single tRNA molecule. The tRNA interacts with three of the four domains of a single AlaRS subunit in an unprecedented orientation. The enzyme's carboxy-terminal domain binds $tRNA^{Ala}$/GU through interactions with the 'elbow' region of the L-shaped tRNA, and the aminoacylation domain interacts with the 3′-CCA end of the tRNA (a single-stranded region at the 3′ terminus that includes the point of attachment for alanine).

Duplex regions of RNA molecules, such as the acceptor stem of $tRNA^{Ala}$, contain two grooves known as the major and minor

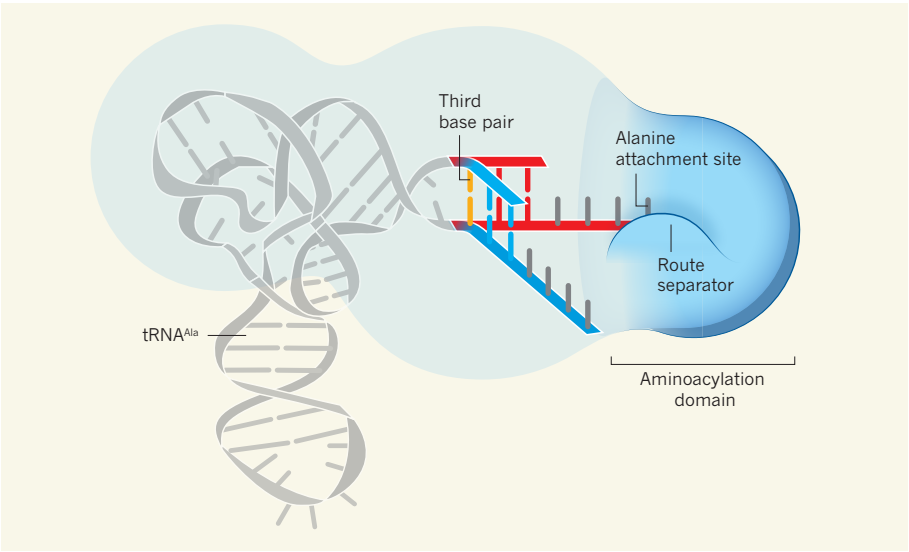


Figure 1 | Conformational effects of a single base pair in a transfer RNA. The AlaRS enzyme catalyses the attachment (aminoacylation) of the amino acid alanine to its matching tRNA ($tRNA^{Ala}$). Naganuma *et al.*[1] report crystal structures of AlaRS in complex with wild-type $tRNA^{Ala}$ (which contains a 'wobble' base pair in the third position of the acceptor-stem helix), and of AlaRS in complex with a mutant $tRNA^{Ala}$ that contains a Watson–Crick base pair instead of the wobble pair. In this cartoon, shaded regions indicate domains of AlaRS that contact the tRNA. The authors find that the acceptor-stem region of $tRNA^{Ala}$ adopts a straight conformation (red) when the wobble base pair is present, but a bent conformation (blue) with the Watson–Crick base pair. Only the straight conformation places the single-stranded region of the RNA into the active site of AlaRS's aminoacylation domain. Amino-acid residues in the aminoacylation domain act as a 'route separator' that prevents transition between the two conformations.

© 2014 Macmillan Publishers Limited. All rights reserved

grooves. Naganuma *et al.* report that recognition of the acceptor stem of $tRNA^{Ala}$/GU is mediated by the tips of two α-helices (α11 and α14, located in the tRNA-recognition domain of AlaRS) that hold the acceptor stem tightly and position it for aminoacylation by clamping onto both the minor and major grooves. The strong clamp formed by α11 and α14 widens the major groove in the acceptor stem, preventing structural clashes with other AlaRS domains.

The researchers also find that the G3·U70 base pair is in the wobble geometry by which G3 orients to the minor groove. The amino group of G3 seems to form two hydrogen bonds with an aspartate amino-acid residue (Asp 450) in α14, which supports the idea[9] that this group has a crucial role in aminoacylation by AlaRS. In contrast to other ARS·tRNA complexes solved so far[13], the L-shaped structure of $tRNA^{Ala}$ in the complex is only mildly distorted compared with that of the free tRNA.

Overall, Naganuma and colleagues' structures show that AlaRS accommodates $tRNA^{Ala}$/AU and $tRNA^{Ala}$/GU in similar ways, but they reveal a previously unknown mechanism by which the local geometry of the G3·U70 wobble pair is propagated through the acceptor-stem duplex towards the single-stranded 3′-CCA region. In the GU-containing complex, the terminus of the 3′-CCA end — the site of amino-acid attachment — reaches the active site of AlaRS (Fig. 1). But in the AU variant, the 3′-CCA end bends away from the active site. This bent conformation might be stabilized by interactions between nearby nucleotides. Even though $tRNA^{Ala}$/AU can, in principle, switch between extended and bent conformations, such a conformational change is restricted by a 'route separator' in AlaRS, a structure formed by three consecutive glycine amino-acid residues and a glutamate residue.

The authors' kinetic analyses confirmed that the k_{cat} value of the *A. fulgidus* AlaRS dimer for aminoacylation of $tRNA^{Ala}$/GU is about 100-fold higher than for $tRNA^{Ala}$/AU, whereas the rates of overall product formation and binding affinities of the tRNA variants differed by only about twofold. Thus, tRNA discrimination by AlaRS seems to rely on a catalytic mechanism involving 'reactive' (straight acceptor stem) and 'non-reactive' (bent acceptor stem) states of the bound tRNA.

Although Naganuma and colleagues' findings provide exciting information about the mechanism by which AlaRS discriminates between $tRNA^{Ala}$/GU and $tRNA^{Ala}$/AU, higher-resolution structures are needed to assign specific hydrogen-bonding interactions with greater confidence. Moreover, we still do not know how AlaRSs recognize mitochondrial $tRNA^{Ala}$ molecules, which do not usually contain a G3·U70 base pair. And it is not clear how the 3′-CCA end of $tRNA^{Ala}$/GU interacts with the second catalytic site (the editing domain) of AlaRS, which hydrolyses tRNAs to which the wrong amino acids have been attached and which also has to recognize the G3·U70 base pair[14].

Crystal structures of other ARSs bound to mutant tRNAs will be needed to establish whether those enzymes also use reactive and unreactive states, as well as route separators, to decode their distinct identity sets. Finally, it remains to be seen whether similar mechanisms are employed by other RNAs that use the G·U wobble pair as a specific recognition signal. ■

Oscar Vargas-Rodriguez *and* **Karin Musier-Forsyth** *are in the Department of Chemistry and Biochemistry, Center for RNA Biology, Ohio State University, Columbus, Ohio 43210, USA.*
e-mail: musier@chemistry.ohio-state.edu

1. Naganuma, M. *et al. Nature* **510,** 507–511 (2014).
2. Hou, Y.-M. & Schimmel, P. *Nature* **333,** 140–145 (1988).
3. McClain, W. H. & Foss, K. *Science* **240,** 793–796 (1988).
4. Giegé, R., Sissler, M. & Florentz, C. *Nucleic Acids Res.* **26,** 5017–5035 (1998).
5. Beuning, P. J. & Musier-Forsyth, K. *Biopolymers* **52,** 1–28 (1999).
6. Imura, N., Weiss, G. B. & Chambers, R. W. *Nature* **222,** 1147–1148 (1969).
7. de Duve, C. *Nature* **333,** 117–118 (1988).
8. Park, S. J., Hou, Y.-M. & Schimmel, P. *Biochemistry* **28,** 2740–2746 (1989).
9. Musier-Forsyth, K. *et al. Science* **253,** 784–786 (1991).
10. Musier-Forsyth, K. & Schimmel, P. *Nature* **357,** 513–515 (1992).
11. McClain, W. H., Chen, Y.-M., Foss, K. & Schneider, J. *Science* **242,** 1681–1684 (1988).
12. Gabriel, K., Schneider, J. & McClain, W. H. *Science* **271,** 195–197 (1996).
13. Perona, J. J. & Hadd, A. *Biochemistry* **51,** 8705–8729 (2012).
14. Beebe, K., Mock, M., Merriman, E. & Schimmel, P. *Nature* **451,** 90–93 (2008).

This article was published online on 11 June 2014.

APPLIED PHYSICS

Trawling for complements

A method has been invented for determining nanoscale variations in the distribution of electric charge on surfaces. It has so far been used to examine specific inorganic materials, but could find widespread applications in imaging.

J. MARTY GREGG & AMIT KUMAR

Thunderstorms are among nature's most awe-inspiring spectacles. Lightning flashes pierce the gloom, causing vast amounts of electric charge — built up at the base of clouds — to find their way to Earth. Even without flashes, static electricity in the clouds affects the charge distribution on Earth's surface. Localized regions of land, and all the objects contained in them, become either enriched in or denuded of electrons, producing a localized net charge that is opposite in sign to that of the clouds directly above. Although most of the cloud base is negative, charge densities vary. As the storm passes overhead, these variations are reflected by complementary charge variations on Earth. To facilitate charge redistribution, electric currents must flow. So, by monitoring the currents at fixed locations on Earth, variations in the charge densities of the passing storm clouds can be mapped. Writing in *Proceedings of the National Academy of Sciences*, Hong *et al.*[1] report how they have done almost exactly this, except that, rather than mapping large-scale electrical storms, they have invented an analogous technique to determine nanoscale charge variations on the surfaces of inorganic materials.

To do their experiments, Hong and colleagues made relatively minor adaptations to an atomic force microscope (AFM), if anything, simplifying its mode of operation over that conventionally used. They brought a sharp conducting tip, positioned on the end of a flexible cantilever, into contact with the surface of a lithium niobate crystal. They then trawled this metallic tip across the crystal surface and, using a sensitive detection circuit, monitored minute currents associated with charge redistribution between the tip and its grounded electrode.

Lithium niobate is an insulating oxide, and is important in telecommunications applications. Within each of the unit cells of this crystalline material, ions are naturally arranged such that the centres of positive and negative charge are slightly separated spatially. This results in what is termed an electric dipole, a vector entity that points from negative to positive within the crystal. In lithium niobate, the orientation of this dipole can be reversed by an electric field, defining the material as ferroelectric. For practical applications, reversals in dipole orientation induced by an electric field are engineered to make spatially periodic stripe patterns, in which positive and negative charges are alternately closer to the crystal surface, in distinct regions called domains. When electrodes are brought into

© 2014 Macmillan Publishers Limited. All rights reserved

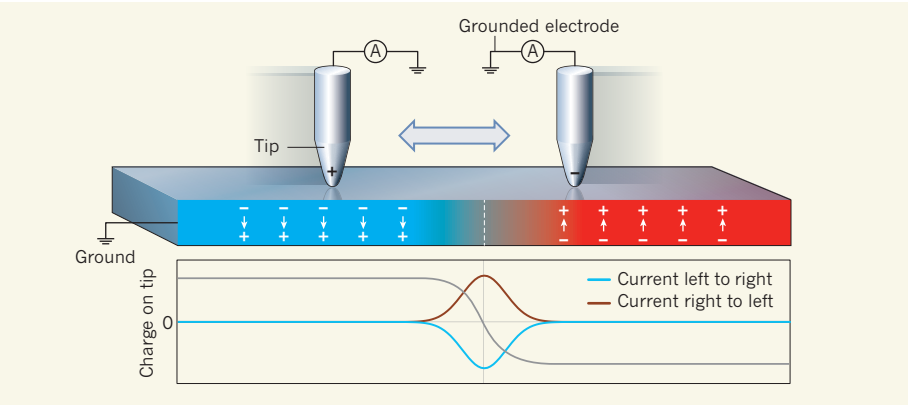


Figure 1 | Charge gradient mapping. Hong and colleagues[1] have developed a technique for determining charge variations on the surface of ferroelectric materials. A metallic tip connected to the ground by means of an electrode scans rapidly across the surface of the material, which is also connected to the ground. The tip records a current, in a device equivalent to an ammeter (A), only when it encounters a charge gradient at a junction between two distinct domains, in which positive and negative charges are alternately closer to the surface; white arrows denote the electric-dipole vectors in each domain. The magnitude of the current (blue and brown curves), which is associated with charge redistribution in the tip (grey curve), is proportional to the charge gradient as well as to the tip velocity; its sign is governed by the direction of the charge gradient, as well as by the scan direction.

contact, the proximity of the near-surface charges in the ferroelectric material induces charge redistribution in the electrode: charges of opposite sign to the near-surface charges in the material are attracted to the interface between the electrode and the ferroelectric material, and like charges are repelled.

The same thing happens when an AFM tip is used as an electrode (Fig. 1). If the tip is moved within a single domain, because dipoles are uniformly oriented, the initial charge redistribution that occurs in the tip is unchanged. However, whenever the tip passes over the junction between one domain and the next (over a domain wall), dipole reversal causes complementary charges to flow between the tip and its grounded electrode. Because the current associated with this charge redistribution between the tip and the ground is proportional to the rate of change of charge at the tip, it increases with the scanning speed of the AFM tip (complete charge redistribution is forced to occur in a shorter time). Moreover, because the nature of the charge redistribution in the tip depends on the direction in which a specific domain wall is traversed, the sign of the current changes whenever the scan direction is reversed.

Hong *et al.* point out that the total current that is generated as the tip passes over each domain wall equates to a charge-density change at the ferroelectric material's surface that is around twice that of the dipole density in lithium niobate, exactly as would be expected. The finding confirms that this method can quantitatively, as well as qualitatively, map nanoscale charge-density variations. The authors highlight the possibly confounding influence of ions and dipolar molecules in the air that nullify the electric fields generated by the near-surface charges in the ferroelectric material. However, they note that their imaging technique, which they call charge gradient microscopy, is largely immune to such airborne layers, because it seems that the tip scrapes them away.

Although this method for imaging charge variation is conceptually quite simple, its potential importance cannot be overstated. It has become evident that the domain walls of ferroelectric materials can have diverse functional characteristics that are completely different from the domains that they delineate: they can be conducting[2], or even superconducting[3], when the domains themselves are insulating. Moreover, they can be moved, injected and annihilated by external electric fields[4], and so could be used in new two-dimensional electronic devices. Hong and colleagues' imaging technique rapidly locates ferroelectric-material domain walls and could therefore be invaluable in supporting the quest for "domain wall nanoelectronics"[5]. Furthermore, its discovery comes at a time when shortcomings in the most common method for domain imaging in ferroelectric materials (piezoresponse force microscopy) are starting to become uncomfortably apparent: for example, high-quality images of 'domains' have been obtained using piezoresponse force microscopy in materials in which ferroelectricity is extremely unlikely[6,7].

Restricting the authors' charge gradient microscopy to imaging ferroelectric domain walls, exciting as this is, would probably be a mistake. Variation in nanoscale charge distribution must be crucial in many aspects of science, from understanding communications in cell biology to teasing out the details of battery electrochemistry. Perhaps the adoption of this imaging technique by other areas of science will allow its full potential to be realized. ■

J. Marty Gregg *and* **Amit Kumar** *are in the Centre for Nanostructured Media, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK.*
e-mails: m.gregg@qub.ac.uk; a.kumar@qub.ac.uk

1. Hong, S. *et al. Proc. Natl Acad. Sci. USA* **111,** 6566–6569 (2014).
2. Sluka, T., Tagantsev, A. K., Bednyakov, P. & Setter, N. *Nature Commun.* **4,** 1808 (2013).
3. Aird, A. & Salje, E. K. H. *J. Phys. Condens. Matter* **10,** L377–L380 (1998).
4. Whyte, J. R. *et al. Adv. Mater.* **26,** 293–298 (2014).
5. Catalan, G., Seidel, J., Ramesh, R. & Scott, J. F. *Rev. Mod. Phys.* **84,** 119–156 (2012).
6. Bark, C. W. *et al. Nano Lett.* **12,** 1765–1771 (2012).
7. Kim, Y. *et al. ACS Nano* **6,** 7026–7033 (2012).

STRUCTURAL BIOLOGY

Enzyme assembly line pictured

Many enzymes form 'assembly lines' containing a series of catalytic modules. Visualization of how the structure of a module shifts during catalysis provides a clearer idea of how such enzymes work. SEE ARTICLE P.512 & LETTER P.560

PETER F. LEADLAY

Living cells contain many molecular machines that integrate and orchestrate essential processes. One of the most familiar is the ribosome, which polymerizes amino acids to form proteins. In the early 1990s, a radically different type of polymerizing assembly line was discovered[1,2]: polyketide synthase (PKS) multi-enzymes, which make polyketide compounds, including many important antibiotics. Remarkably, these complexes use a different module of catalytic sites for each cycle of a chemical-chain extension. Efforts to understand these sophisticated catalysts have received a great boost from two studies in this issue from the same research group: Dutta *et al.*[3] (page 512) report the three-dimensional structure of a PKS protein that houses an intact catalytic module, and

© 2014 Macmillan Publishers Limited. All rights reserved

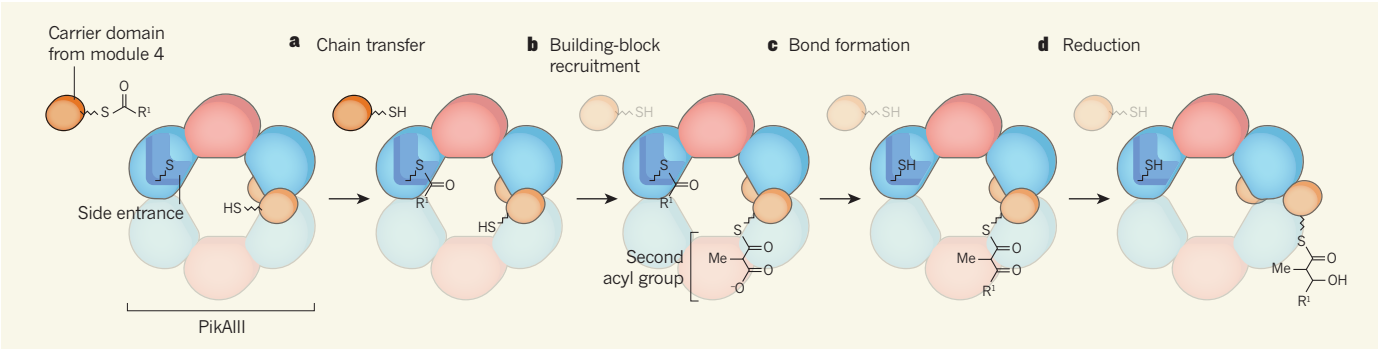


Figure 1 | Close-up of a modular assembly-line enzyme. PikAIII is the fifth module of six in the multi-enzyme that constructs the antibiotic pikromycin. It contains domains that perform substrate recruitment (red), substrate carriage (orange) and catalysis (blue). Dutta *et al.*[3] report that PikAIII is a homodimer that defines an internal reaction chamber (for clarity, only one subunit is shown in full colour; the other is shown in outline). Whicher *et al.*[4] suggest how the domains move as PikAIII catalyses a chain-extension cycle. **a**, An acyl group (COR^1) is transferred from a carrier domain in module 4 to module 5 through a 'side entrance'. R^1 represents a chain containing five molecular building blocks. **b**, The carrier domain of module 5 recruits another acyl building block. Me, methyl group. **c**, **d**, It then moves between the catalytic sites to enable a bond-forming reaction between the two acyl groups and the reduction of a carbonyl group (C=O) to a hydroxyl group (OH), before flipping out of the reaction chamber to present the extended chain to module 6.

Whicher and colleagues[4] (page 560) investigate how individual domains move during the extension cycle.

Biochemical evidence[5] suggests that modular PKS multi-enzymes are homodimeric (composed of two identical subunits), and that the subunits are aligned head-to-head, tail-to-tail and are tightly wound around each other. A functional module thus contains two copies of each type of domain — just as in animal fatty-acid synthase, a complex that uses a very similar set of enzymes to those in a PKS module. In 2008, the crystal structure of animal fatty-acid synthase was obtained[6], revealing an X-shaped molecule in which several of the domains are highly mobile. An independent study[7] reconstructed the three-dimensional structure of this enzyme from single-particle electron micrographs. This confirmed that animal fatty-acid synthase is exceptionally flexible and that sets of domains change position in response to substrate binding, hinting that PKS modules might behave like this too. Many excellent studies have provided static X-ray crystal structures of individual domain fragments of PKS modules, but Dutta *et al.* and Whicher *et al.* are the first to visualize an intact homodimeric module in sufficient detail to reveal its architecture and the dynamic changes that occur during the chain-extension cycle.

The researchers studied PikAIII, a PKS protein that forms module 5 of the bacterial assembly line that makes the antibiotic pikromycin. The extension cycle in PikAIII starts when the protein accepts a partially built polyketide chain (containing an acyl chemical group) from the carrier domain of the previous module, and recruits a new building block (another acyl group) to its own carrier domain (Fig. 1). The tethered acyl groups are then bonded together in a reaction catalysed by a ketosynthase domain, forming an extended chain that ends up on PikAIII's carrier protein. An additional processing domain catalyses a reduction reaction of the extended chain, which is finally handed on to the next module.

Dutta and co-workers used electron cryo-microscopy (cryo-EM) to obtain three-dimensional reconstructions of PikAIII at a resolution of 7 to 10 ångströms. This is not high enough to resolve individual atoms, but by fitting known crystal structures of individual domains into place, the authors were able to visualize the module in its entirety. They found that the identical subunits were indeed aligned head-to-head and tail-to-tail, as predicted[5]. The subunits enclose a single inner chamber into which all the active sites of the domains face. This arrangement is different from that found for animal fatty-acid synthase, and from that recently proposed for PKSs on the basis of low-resolution X-ray scattering analysis[8], but it helps to explain how PKSs control the sequence of events. The authors also observed that the carrier proteins are mobile and may separate from each other so that the molecule seems to adopt a horseshoe shape.

Next, Dutta and colleagues made a construct in which PikAIII was covalently tethered to the carrier protein of the previous module, with the carrier bearing a synthetic polyketide acyl group. Cryo-EM analysis revealed that two different channels provide access to the catalytic site of the bond-forming ketosynthase domain. The externally delivered acyl group gains access through a 'side entrance', whereas the PikAIII carrier protein offers the new building block for bond formation through an opening from the internal chamber (Fig. 1).

Whicher *et al.* used cryo-EM to study PikAIII samples in which a synthetic acyl group was directly attached either to the ketosynthase domain or to the PikAIII carrier protein, to mimic intermediates in the extension cycle. In each case, significant shifts were noted in the position of the carrier protein and the other domains, compared to Dutta and colleagues' structure of non-acylated PikAIII. It therefore seems that the acyl chain is specifically recognized by the catalytic domains, which helps to explain why naturally occurring PKSs normally yield a single product. A similar conclusion has been reached for fatty-acid synthase in recent work[9] that used chemical probes to trap carrier domains as they dock to catalytic domains. Deeper understanding of this aspect of PKS selectivity should help synthetic biologists to redesign these assembly-line synthases to prepare potentially useful analogues of naturally occurring antibiotics.

A key issue that cannot be fully addressed by structural studies at this level of resolution is whether some mechanism actively directs the carrier protein and its cargo to the next 'correct' reaction partner, or whether they simply find their way by diffusion within the chamber. Also unclear is whether the loaded carrier protein can reach all catalytic active sites on either subunit, or is confined to one subset. Likewise, what prevents the extended chain from being prematurely passed to the next module before all the programmed operations in module 5 have taken place? Addressing these questions will require further genetic, biochemical and biophysical work. Nonetheless, by showing that an intact module is much more than the sum of its parts, this elegant study has given fresh impetus to our search for the answers. ■

Peter F. Leadlay *is in the Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK.*
e-mail: leadlay@mole.bio.cam.ac.uk

1. Cortés, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. *Nature* **348,** 176–178 (1990).
2. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. *Science* **252,** 675–679 (1991).
3. Dutta, S. *et al. Nature* **510,** 512–517 (2014).
4. Whicher, J. R. *et al. Nature* **510,** 560–564 (2014).
5. Staunton, J. *et al. Nature Struct. Mol. Biol.* **3,** 188–192 (1996).
6. Maier, T., Leibundgut, M. & Ban, N. *Science* **321,** 1315–1322 (2008).
7. Brignole, E. J., Smith, S. & Asturias, F. J. *Nature Struct. Mol. Biol.* **16,** 190–197 (2009).
8. Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. *J. Mol. Biol.* **426,** 2229–2245 (2014).
9. Nguyen, C. *et al. Nature* **505,** 427–431 (2014).

This article was published online on 18 June 2014.

© 2014 Macmillan Publishers Limited. All rights reserved

REVIEW

doi:10.1038/nature13384

An overview of N-heterocyclic carbenes

Matthew N. Hopkinson[1], Christian Richter[1], Michael Schedler[1] & Frank Glorius[1]

The successful isolation and characterization of an N-heterocyclic carbene in 1991 opened up a new class of organic compounds for investigation. From these beginnings as academic curiosities, N-heterocyclic carbenes today rank among the most powerful tools in organic chemistry, with numerous applications in commercially important processes. Here we provide a concise overview of N-heterocyclic carbenes in modern chemistry, summarizing their general properties and uses and highlighting how these features are being exploited in a selection of pioneering recent studies.

Defined as neutral compounds containing a divalent carbon atom with a six-electron valence shell, carbenes are an intriguing class of carbon-containing compounds. Their incomplete electron octet and coordinative unsaturation, however, render free carbenes inherently unstable and they have been traditionally considered only as highly reactive transient intermediates in organic transformations such as cyclopropanation. Despite attempted syntheses from as early as 1835 (ref. 1), the isolation and unambiguous characterization of a free, uncoordinated carbene remained elusive until pioneering studies in the late 1980s and early 1990s (ref. 2). In a seminal publication in 1988, Bertrand and co-workers reported the preparation of the first isolable carbene stabilized by favourable interactions with adjacent phosphorus and silicon substituents[3]. Three years later Arduengo *et al.* reported an isolable and 'bottleable' carbene incorporated into a nitrogen heterocycle[4]. With structural features inspired by earlier insightful studies by Wanzlick[5] and Öfele[6] on metal–carbene complexes, the remarkable stability and relatively simple synthesis of the first N-heterocyclic carbene (NHC) 1,3-di(adamantyl)imidazol-2-ylidene (IAd, compound labelled **1a**) led to an explosion of experimental and theoretical studies with libraries of novel NHCs being synthesized and analysed (see below). As a result of these investigations, NHCs have been elevated from mere laboratory curiosities to compounds of enormous practical significance as more and more of the rich chemistry of these compounds has been revealed and exploited. As excellent ligands for transition metals, NHCs have found multiple applications in some of the most important catalytic transformations in the chemical industry, while their reactivity upon coordination to main group elements and as organocatalysts has opened up new areas of research.

In this review, we aim to provide a concise overview of the properties and broad range of applications of NHCs, which we hope will serve as a useful introduction and reference guide for scientists interested in studying and applying these important compounds. After an initial summary of the general structure and properties of NHCs, the reactivity and applications in modern chemistry are loosely categorized into three sections with a discussion of NHCs as ligands for transition metals, upon coordination to *p*-block elements and as organocatalysts. Each section contains a brief overview of the key features and major applications with references to seminal publications and more comprehensive specialized reviews provided for more in-depth reading. The discussion is interspersed with more detailed descriptions of selected recent studies, which demonstrate the current state of the art and future trends as an ever-increasing number of NHCs continue to find new and exciting applications across the chemical sciences (Fig. 1).

Structure and general properties of NHCs

In this review, NHCs are defined as heterocyclic species containing a carbene carbon and at least one nitrogen atom within the ring structure[7,8]. Within these criteria fall many different classes of carbene compounds with various substitution patterns, ring sizes and degrees of heteroatom stabilization[9]. A representation of the general structures of NHCs, as exemplified for the first reported compound IAd (**1a**), is shown in Fig. 2a. The overall electronic and steric effects of these structural features go some way to explaining the remarkable stability of the carbene centre C^2. As demonstrated in IAd by the two adamantyl groups bound to the nitrogen atoms, NHCs generally feature bulky substituents adjacent to the carbene carbon, which help to kinetically stabilize the species by sterically disfavouring dimerization to the corresponding olefin (the Wanzlick equilibrium). The electronic stabilization provided by the nitrogen atoms, however, is a much more important factor. In contrast to classical carbenes, NHCs such as IAd exhibit a singlet ground-state electronic configuration with the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) best described as a formally sp^2-hybridized lone pair and an unoccupied *p*-orbital at the C^2 carbon, respectively (Fig. 2b). The adjacent σ-electron-withdrawing and π-electron-donating nitrogen atoms stabilize this structure both inductively by lowering the energy of the occupied σ-orbital and mesomerically by donating electron density into the empty *p*-orbital. The cyclic nature of NHCs also helps to favour the singlet state by forcing the carbene carbon into a bent, more sp^2-like arrangement. This ground-state structure is reflected in the C^2-N bond lengths (1.37 Å) observed in IAd, which fall in between those of its corresponding imidazolium salt ($IAdH^+$, 1.33 Å)[4] and its C^2-saturated analogue ($IAdH_2$, 1.49 Å)[10], signifying that the C^2-nitrogen bonds possess partial double-bond character.

These general principles of carbene stabilization apply to all classes of NHC although the relative importance of each effect varies from compound

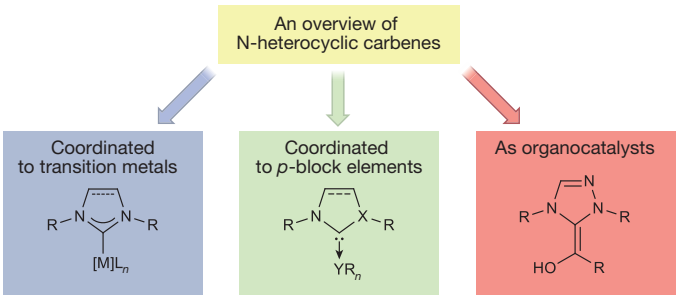


Figure 1 | Major applications of NHCs. In this review, the major applications of NHCs in modern chemistry are divided into the three categories shown. In the first two, the features and uses of NHCs coordinated to metals and *p*-block elements are discussed, while the third section details the applications of NHCs as organocatalysts.

[1]Organisch-Chemisches Institut, Westfälische Wilhelms-Universität Münster, Corrensstrasse 40, 48149 Münster, Germany.

©2014 Macmillan Publishers Limited. All rights reserved

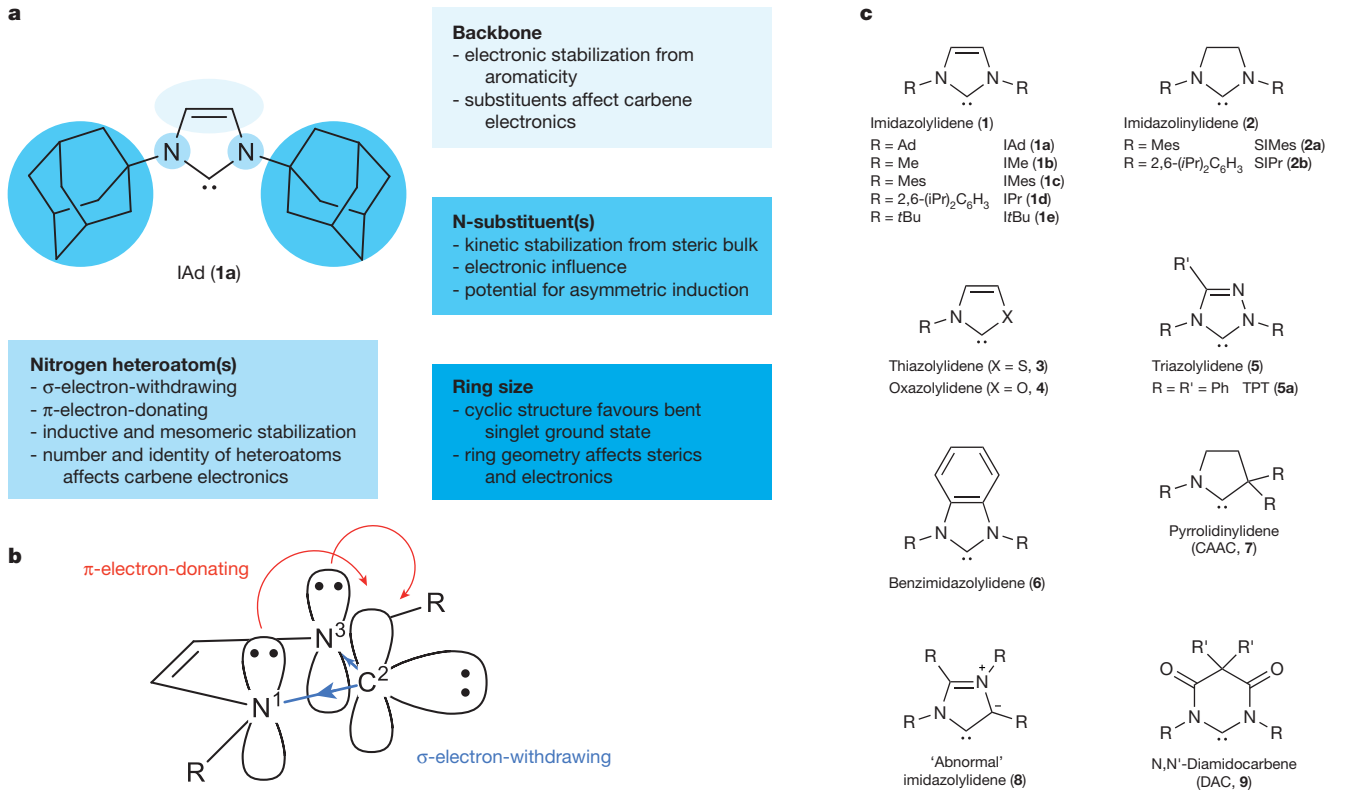


Figure 2 | Structural features of NHCs. **a**, General structural features of IAd (**1a**), detailing the effects of the ring size, nitrogen heteroatoms and the ring backbone and nitrogen-substituents on the stability and reactivity of the NHC. **b**, Ground-state electronic structure of imidazol-2-ylidenes. The σ-withdrawing and π-donating effects of the nitrogen heteroatoms help to stabilize the singlet carbene structure. **c**, Structures of some of the most commonly applied classes of NHCs. Ad, adamantyl; Mes, mesityl; tBu, *tert*-Butyl; iPr, *iso*-propyl; Ph, phenyl.

to compound (Fig. 2c). NHCs derived from heteroaromatic compounds benefit from a greater degree of stabilization by virtue of their partial aromaticity. This effect, which has been calculated to be around 25 kcal mol^{-1} for model imidazol-2-ylidenes[11], allows for a lesser demand for proximal steric bulk and, consequently, the simple methyl-substituted NHC 1,3-di(methyl)imidazol-2-ylidene (IMe, compound labelled **1b**) is persistent in solution[12]. There are, however, many stable carbenes that do not benefit from aromaticity, with the first example, 1,3-di(mesityl)imidazolin-2-ylidene (SIMes, **2a**), being reported by Arduengo and co-workers in 1995[13]. Neither is there a requirement for two adjacent nitrogen atoms to stabilize the carbene centre[14]. NHCs bearing alternative heteroatoms such as sulphur (**3**) and oxygen (**4**) are accessible, while stable carbenes containing only one nitrogen substituent, such as the series of cyclic (alkyl)(amino)carbenes (CAACs, **7**) introduced by Bertrand *et al.*[15], have also received considerable research attention.

Similar species stabilized by only one nitrogen atom may be formed upon generation of the carbene centre at alternative positions to C^2. These mesoionic or 'abnormal' carbenes **8**, for which a neutral, non-zwitterionic carbene resonance structure cannot be drawn, are generally more electron-donating than their 'normal' analogues and can display very different properties[16,17]. Remote NHCs, where the carbene carbon is not situated adjacent to a nitrogen heteroatom have also been reported. The size and substitution pattern of the nitrogen heterocycle can have a large effect on the properties of the carbene. While 5-membered rings still make up the largest class of NHCs, examples containing smaller or larger ring sizes including N,N′-diamidocarbenes (DACs, **9**) have also been reported. These latter compounds lead to increased steric shielding owing to the greater $N^1-C^2-N^3$ bond angle, which effectively pushes the nitrogen substituents closer to the carbene centre. Larger ring sizes also have an electronic effect as geometric constraints imposed by the cyclic structure cause variations in the degree and nature of heteroatom stabilization. It is also worth noting that several related classes of stable carbenes are known, which, although not defined as NHCs, benefit from similar modes of stabilization. These include acyclic derivatives and cyclic species featuring different ring heteroatoms such as phosphorus instead of nitrogen[7,14,17].

The ground-state electronic structure of NHCs provides a framework for understanding their reactivity. In contrast to the typical electrophilicity of most transient carbenes, the lone pair situated in the plane of the heterocyclic ring of NHCs renders these compounds nucleophilic. The principal consequence of this characteristic is the propensity of NHCs to act as σ-donors and bind to a wide range of metallic and non-metallic species. The extraordinary strength and distinct features of these interactions and their influence on the stability, structure and reactivity of the resulting complexes or adducts forms the basis for the meteoric rise in interest of NHCs. This extensive coordination chemistry and the various applications of NHCs arising from it are discussed in more detail in the following sections of this review.

Another attractive feature of NHCs is the comparative ease with which libraries of structurally diverse analogues can be prepared and studied. In the majority of cases, the carbene is generated upon deprotonation of the corresponding cationic heterocyclic azolium salt and, as a result, synthetic routes to NHCs benefit from centuries of research on the preparation of heterocyclic compounds[18]. For most classes of N-heterocycles, simple variation of the starting materials in a modular synthetic sequence allows for facile modification of the steric and electronic properties of the resulting carbene. The nitrogen-substituents or other groups situated adjacent to C^2 have the largest influence on the steric environment at the carbene centre with different classes of heterocycle having inherently different steric requirements. The NHC electronics are primarily governed by the class of heterocycle, with the substitution pattern of the ring backbone also playing an important part. The quantification of these properties facilitates easy comparison both between different NHCs and between NHCs and other related compounds such as phosphines, and allows for more enlightened selection of the appropriate carbene for any given application[19,20].

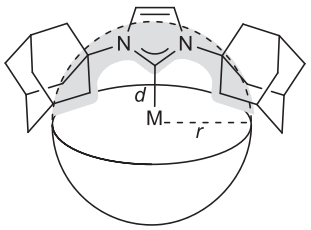
©2014 Macmillan Publishers Limited. All rights reserved

BOX 1

Quantitative measures of steric and electronic properties

The steric properties of NHCs can be conveniently quantified using the 'buried volume' parameter ($\%V_{bur}$) developed by Nolan, Cavallo and co-workers[21]. As shown in Box 1 Fig. 1 (in which M indicates metal and N indicates nitrogen), the $\%V_{bur}$ value of an NHC refers to the percentage of a sphere occupied or 'buried' by the ligand upon coordination to a metal at the centre of the sphere. Fixed parameters of 2 Å for the metal–carbene bond distance *d* and 3 Å or 3.5 Å for the sphere radius *r* are typically used with a larger $\%V_{bur}$ value signifying a greater steric influence of the ligand on the metal centre. The buried volume can be determined from crystallographic data or from theoretical calculations with the free NHC, various NHC–metal complexes or the azolium salt precursor being suitable data sources. The derived $\%V_{bur}$ values, however, may vary greatly depending on the system used and care must be taken to compare only values determined using the same approach.

The electronic properties of NHCs are most commonly described using the Tolman electronic parameter (TEP)[22]. Originally developed for phosphines, the TEP specifically evaluates the electron-donating ability of a ligand (L) by measuring the infrared-stretching frequencies of carbonyl ligands in model transition metal carbonyl complexes. The more electron-donating the ligand of interest the more electron-rich the metal centre becomes, increasing the degree of π-backbonding into the carbonyl ligands and thus reducing their bond order and infrared stretching frequency. Although $[LNi(CO)_3]$ complexes were initially the model species for TEP calculation, the less toxic complexes *cis*-$[LIrCl(CO)_2]$ and *cis*-$[LRhCl(CO)_2]$ are nowadays more prevalent. Mathematical formulae have been derived to correlate the TEP values obtained from different complexes. Significant variation in TEP values may arise, however, depending on the resolution of the infrared spectrometer (TEP values for NHCs span only about 10 cm^{-1}) and the solvent used (typically CH_2Cl_2)[19,20].



Two of the most commonly measured parameters; the buried volume ($\%V_{bur}$) for sterics[21], and the Tolman electronic parameter (TEP) for electronics[22] are outlined in Box 1. The diversity and modular nature of synthetic approaches also allow for easy incorporation of chiral information into the NHC structure. Libraries of derivatives have been prepared, featuring enantiomerically enriched moieties either at the backbone or at the nitrogen substituents.

Coordination of NHCs to transition metals

The majority of applications of N-heterocyclic carbenes involve their coordination to transition metals (Fig. 3). The first examples of NHC–metal complexes predate the isolation of a free NHC by over 20 years, with Wanzlick[5] and Öfele[6] independently synthesizing imidazol-2-ylidene-bearing mercury(II) and chromium(0) species, respectively, in 1968. Insightful studies before the isolation of IAd were also conducted by Lappert and co-workers during the early 1970s (ref. 23). As mentioned above, the suitability of NHCs as ligands for transition metals can be rationalized by their inherent σ-donor ability with a formal sp^2-hybridized lone pair available for donation into a σ-accepting orbital of the transition metal. The full nature of bonding in these complexes has been studied by a number of different groups and has been the subject of reviews by Díez-González and Nolan[24], and Cavallo and co-workers[25]. While σ-donation is the most important component of metal–ligand binding, the contribution of both π-back-bonding into the carbene *p*-orbital and π-donation from the carbene *p*-orbital may not be inconsiderable. For example, Frenking and co-workers calculated that π-contributions account for about 20% of the overall bond energy in group-11 metal-imidazol-2-ylidene and imidazolin-2-ylidene complexes[26]. However, in practice, metal–C^2 coordination is generally drawn as a single rather than a double bond with π-contributions restricted to delocalization within the NHC ring (often depicted by a curved line between the ring heteroatoms). This representation best reflects the experimentally observed potential for rotation around the metal–C^2 bond and emphasizes the differences between NHCs and conventional Fischer or Schrock carbene ligands.

The strong σ-donor and comparatively weak π-acceptor properties of NHCs bear similarities to the coordination characteristics of phosphines, and they were initially considered as mimics for this pervasive class of ancillary ligand in transition-metal coordination chemistry[27]. There are, however, a number of differences between the two classes of ligands. As indicated by their lower TEP values (see Box 1 for a description of TEP

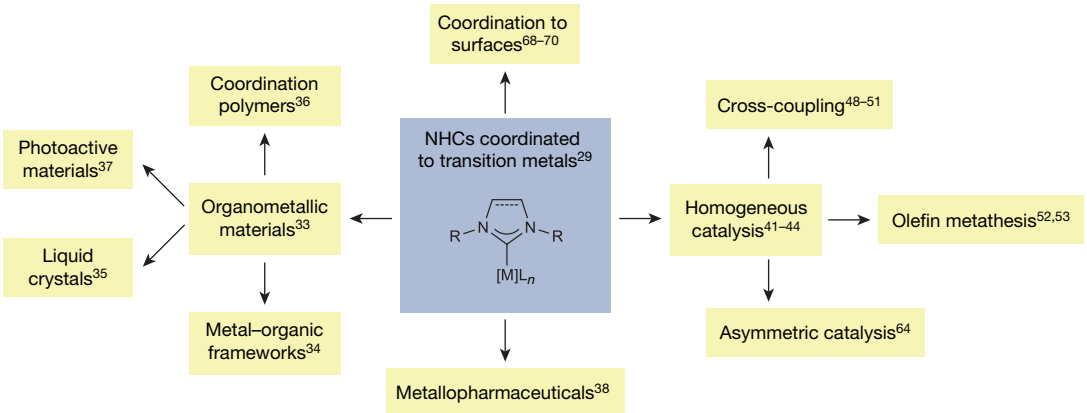


Figure 3 | Major applications of NHCs coordinated to transition metals. NHCs are excellent σ-donors and readily bind transition metals. This feature has led to the most important application of NHCs as ancillary ligands in homogeneous transition-metal catalysis. NHC-Metal complexes also find many different applications as organometallic materials and as metallopharmaceuticals.

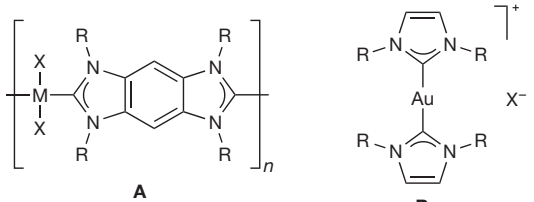
©2014 Macmillan Publishers Limited. All rights reserved

BOX 2

Medicinal and materials applications of NHC–metal species

The high thermal stability of metal–NHC complexes and the ability to tune their steric and electronic properties are attractive features for the development of organometallic materials[33]. Imidazolium salts incorporated as linker molecules in metal–organic frameworks have been successfully coordinated to transition metals to afford organometallic complexes within the pores of the material[34]. Transition-metal complexes bearing NHCs with hydrophobic long alkyl chain N-substituents may also self-assemble to form highly air- and moisture-resistant liquid crystalline materials, which are thermally stable beyond the clearing point[35]. The incorporation of NHC–metal complexes into the side chains or main chain of polymers has been extensively studied. Using benzene-linked bis(NHC) units, Bielawski and co-workers prepared a series of well-defined palladium(II) and platinum(II) organometallic polymers **A** (see Box 2 Fig. 1), which display self-healing properties by virtue of the inherent reversibility of metal–ligand coordination[36]. Materials of this type show promise as electrical conductors with conjugated bis-NHC linkers allowing electronic coupling between the two coordinated metal centres. NHC–transition metal complexes that act as phosphors and other photoactive materials have also been reported[37].

An increasing number of publications have focused on the medicinal applications of NHC–transition-metal complexes as metallopharmaceuticals with silver(I) and gold(I) species showing particular promise as antibacterial and anticancer agents, respectively[38]. Many imidazol-2-ylidene– and imidazolin-2-ylidene–Ag complexes exhibit impressively low minimum inhibitory concentration values against a range of Gram-positive and Gram-negative bacteria ($<10\,\mu g\,ml^{-1}$). In comparison to the standard reference $AgNO_3$, these species are generally therapeutically active for a longer period, which may be rationalized by a slower release of active Ag^+ ions from the NHC-stabilized complexes. NHC–metal species featuring gold have shown promise as anticancer drugs based on targeting of mitochondria. For these systems, the ability to fine-tune the lipophilicity of the complexes through modification of the N-substituents on the NHC is crucial, because anticancer activity is highly dependent on penetration through the mitochondrial membrane. In a seminal publication by Berners-Price, Filipovska and co-workers, the cationic gold(I) complexes **B** (see Box 2 Fig. 1; in which Et indicates ethyl, nPr indicates *n*-propyl and Bn indicates benzyl) were shown to induce apoptosis by selective inhibition of the selenoenzyme thioredoxin reductase, which is overexpressed in many human cancers[39].



A
R = Bu, Bn
M = Pd, Pt
X = Cl, Br

B
R = iPr, nPr, Et
X = Cl, Br

values), NHCs are in general more electron-donating than phosphines. This leads to thermodynamically stronger metal–ligand bonds and is reflected in the typically greater bond dissociation energies and shorter metal–ligand bond lengths observed for NHC complexes over their phosphine counterparts. Notable exceptions to this trend arise, however, when steric constraints interfere with metal–ligand binding. A comparison of relative bond dissociation energy values for a range of imidazol-2-ylidene–ruthenium(II) and imidazolin-2-ylidene–ruthenium(II) complexes ([Cp*Ru(NHC)Cl], where Cp* = C_5Me_5) plotted against the buried volume (%V_{bur}, see Box 1 for a description) shows a marked linear decrease in NHC–Ru bond strength of about 12 kcal mol^{-1} with increasingly bulky NHCs (%V_{bur} from 23–37%)[21]. For all carbene ligands except the most sterically demanding IAd (**1a**), however, the bond dissociation energies remained greater than that of the analogous complex with even the most Lewis-basic phosphine tested (PCy_3). As a rule, the stronger metal–ligand interaction renders NHC–metal coordination less labile than metal–phosphine binding and the complexes are more thermally and oxidatively stable[28]. A comparison of the steric properties of NHCs and phosphines also reveals significant differences. Whereas the sp^3-hybridization of phosphines results in a cone-shaped spatial arrangement of the steric bulk, most classes of NHCs, including the most commonly employed imidazole-derived types **1** and **2**, can best be described as fan- or umbrella-shaped with the nitrogen-substituents adjacent to the carbene carbon oriented more towards the metal. As a result, NHCs are generally considered to be sterically demanding ligands with variations in the nitrogen substituents and class of heterocycle having a large effect on the steric environment at the metal centre. In contrast to phosphines, the steric properties of NHCs are also highly anisotropic and rotation around the metal–carbene bond may occur so as to minimize clashing with other bulky ligands.

A further disparity between NHCs and phosphines concerns the ease of varying their steric and electronic properties. As discussed above, many synthetic routes employing well established heterocyclic chemistry are known for NHCs, whereas structural variation of phosphines is often not trivial. Moreover, in contrast to phosphines, where changing the phosphorus substituents invariably affects both the steric and electronic properties, the potential to separately modify the nitrogen substituents, backbone functionality and class of heterocycle in NHCs allows for more independent variation of each parameter.

With these attractive features, NHCs nowadays rival phosphines and cyclopentadienyls as ligands across organometallic chemistry and the extensive range of accessible complexes continues to grow at an astonishing rate. NHC complexes have been described for all transition metals[29] and in various oxidation states while similar adducts are known for alkali and alkali earth[30], and *f*-block metals[31]. Various methods of synthesizing complexes may be employed and there is generally no requirement to preform the free carbene. Most commonly, *in situ* deprotonation of an azolium salt is conducted in the presence of a suitable transition-metal precursor, although strategies involving α-elimination or oxidative addition at the carbene carbon, carbene transfer from pre-formed NHC-silver(I) or copper(I) complexes and metal-templated construction of the NHC may also be used. Azolium salts are typically bench-stable solids and precursors for the most widely used NHCs, such as IPr (**1d**) and SIMes (**2a**) are commercially available. An interesting class of NHCs incorporates additional tethered coordinating groups into their structure and many examples of bi-, tri- and tetradentate ligands featuring multiple NHC moieties have been reported. These species may either bridge several different metals or act as chelating ligands to a single metal, depending on their structure, affording complexes with a variety of different geometries[32].

The attractive features of NHC–metal coordination have led to a wide range of different applications of these complexes across the chemical sciences. A summary of some of their uses in materials[33–37] and as metallopharmaceuticals[38,39] is provided in Box 2. By far the largest application of NHC–transition metal complexes, and indeed of NHCs in general, however, is in the homogeneous catalysis of organic transformations. First demonstrated by Herrmann and co-workers in an imidazol-2-ylidene-palladium-catalysed Mizoroki–Heck reaction[40], NHC complexes of various transition metals are, nowadays, privileged catalysts for a myriad of academically and commercially important processes. The huge number of

©2014 Macmillan Publishers Limited. All rights reserved

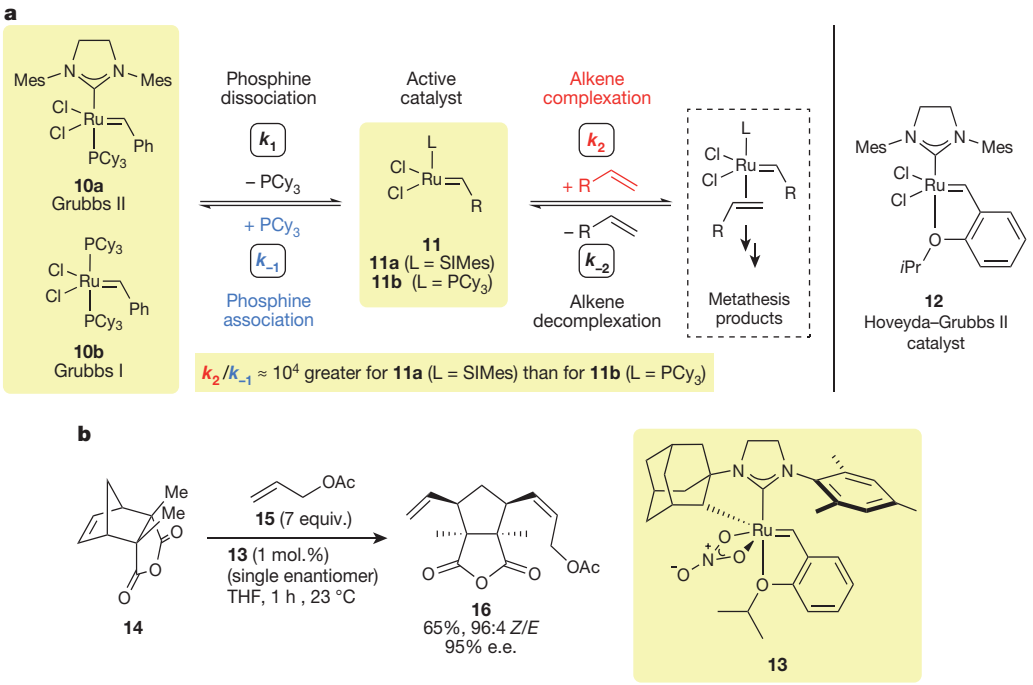


Figure 4 | NHCs as ligands in ruthenium-catalysed olefin metathesis. **a**, Comparison of Grubbs' first- and second-generation catalysts (**10**) in (cross) metathesis reactions. The NHC-bearing Grubbs' second-generation catalyst (**10a**) has an approximately 10^4-fold greater affinity for the alkene substrates, resulting in higher reactivity than the phosphine-coordinated first-generation catalyst **10b** (ref. 54). **b**, *Z*-selective olefin metathesis using a Ru-NHC catalyst recently developed by Grubbs and co-workers[57]. The steric environment in this catalyst is fixed by a key sp^3-C−H activation at the adamantyl group, which leads to the high *Z*-selectivity and, in this case, enantioselectivity with complex **13** being used as a single stereoisomer. e.e., enantiomeric excess.

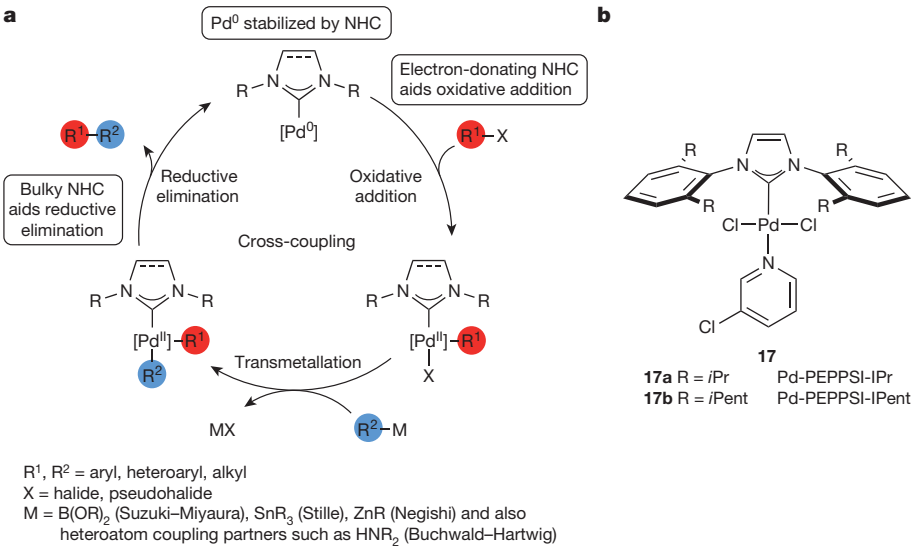
publications on this topic have been excellently covered in several book chapters and review articles and here we provide only a short summary of some of the key features[41–44]. A by no means comprehensive selection of some of the most important transformations mediated by NHC–metal complexes includes Ir- and Ru-catalysed hydrogenation and hydrogen transfer[45], gold-catalysed activation of π-bonds[46] and Rh- and Pt-catalysed hydrosilylation[47]. The two most extensively studied classes of catalytic reactions, however, are cross-coupling (catalysed by palladium or other metals)[48–51] and ruthenium-catalysed olefin metathesis[52,53].

Much of the success of NHC spectator ligands in these transformations can be attributed to the increased catalyst stability, and consequent lower rates of catalyst decomposition resulting from strong metal–ligand binding. The distinct steric and electronic influence of the NHC on the metal centre may also lead to improved catalytic activity. These two factors are elegantly demonstrated in the excellent efficiency of arguably the most famous NHC–metal complex; Grubbs' second-generation olefin metathesis catalyst **10a** (Fig. 4a)[52,53]. In comparison to the first-generation Grubbs' catalyst **10b**, which features two PCy_3 ligands bound to ruthenium, the SIMes–Ru(II) complex **10a** exhibits substantially greater thermal stability and remains catalytically active for cross- and ring-closing metathesis, and ring-opening metathesis polymerization (ROMP) reactions at much lower catalyst loadings. Moreover, the NHC-stabilized complex displays significantly higher reactivity than its predecessor and has dramatically expanded the range of suitable substrates for metathesis reactions. Mechanistic studies have revealed that the difference in activity of the two-generation catalysts results from the relative affinity for the alkene substrates of their respective active catalytic species, formed upon dissociation of PCy_3 (**11a** and **11b**)[54]. Catalytically relevant binding to the π-accepting alkene substrates over the σ-donating PCy_3 (leading back to **10**) was found to be four orders of magnitude more favourable for the more electron-rich NHC-stabilized ruthenium centre in **11a** than for phosphine-coordinated **11b**. Numerous related NHC–ruthenium complexes, including the widely used Hoveyda–Grubbs second-generation catalyst (**12**), have been developed for applications in metathesis and the importance of these kinds of reactions was recognized with the award of the Nobel Prize for chemistry in 2005.

As well as improving the efficiency, selectivity and practicality of transition-metal-mediated reactions, NHC-bearing catalysts can give access to unprecedented reactivity pathways and selectivity paradigms that are not observed using other ligands. Building on the foundations laid over the last two decades, many research programmes are currently focused on the development of the next generation of carbene–metal catalysts employing new NHCs. One such system has been recently developed in a series of groundbreaking publications by the Grubbs group. As a result of the dynamic nature of metathesis reactions, previously developed ruthenium catalysts such as Grubbs' first- and second-generation catalysts (**10**) and the Hoveyda–Grubbs catalyst (**12**) favour formation of the thermodynamic *E*-alkene products. The selective generation of *Z*-configured alkenes, which feature widely in natural products and biologically and industrially relevant compounds, has been a major challenge limiting the usefulness of ruthenium-catalysed olefin metathesis.

The major breakthrough in this field came in 2011 with the preparation of ruthenium(II) complexes bearing an unsymmetrical N-adamantyl-substituted imidazolin-2-ylidene ligand that displayed high *Z*-selectivity in cross-metathesis reactions[55]. Further optimization of the catalyst structure led to the development of κ^2-nitrate-coordinated complex **13**, which has recently shown remarkable *Z*-selectivity in a range of transformations, including ROMP and ring-closing metathesis[56]. The key feature in complex **13** and related catalysts is the presence of a second bond between the metal and the NHC resulting from sp^3-C−H activation at the N-adamantyl substituent. As demonstrated by density functional theory (DFT) calculations, this chelation restricts rotation around the metal–carbene bond and fixes the N-mesityl substituent over the catalytically relevant alkylidene moiety. Steric clashing between these groups in the transition state then explains the *Z*-selectivity. A further development of this methodology was reported in 2013 in the asymmetric ring-opening/cross-metathesis of norbornene derivatives[57]. Chromatographic separation of two intermediate carboxylate complexes followed by salt metathesis with sodium nitrate

©2014 Macmillan Publishers Limited. All rights reserved



R¹, R² = aryl, heteroaryl, alkyl
X = halide, pseudohalide
M = $B(OR)_2$ (Suzuki–Miyaura), SnR_3 (Stille), ZnR (Negishi) and also heteroatom coupling partners such as HNR_2 (Buchwald–Hartwig)

Figure 5 | NHCs as ligands in palladium-catalysed cross-coupling. **a**, General mechanism of palladium-catalysed cross-coupling reactions between organic halides and organometallic reagents. The electron richness of NHCs can lead to improved oxidative addition while reductive elimination can benefit from the steric bulkiness of the NHC. Both steric and electronic factors also help to stabilize the active Pd^0 catalyst. **b**, Structure of Pd-PEPPSI-NHC pre-catalysts (**17**) introduced by Organ and co-workers[48,51] (where PEPPSI is pyridine enhanced precatalyst preparation stabilization). These easy-to-handle complexes are convenient sources of active palladium catalysts for a variety of cross-coupling reactions.

delivered an enantio-enriched sample of complex **13**, which is stereogenic at the ruthenium atom. Reacting 1 mol.% of this complex in the presence of norbornene **14** and excess allyl acetate **15** delivered the corresponding ring-opened product **16**, which features four contiguous and two quaternary stereocentres in 95% enantiomeric excess and in 96:4 selectivity for the *Z*-alkene (yield 65%, Fig. 4b).

A rationale based on improved catalyst stability and reactivity has also been used to explain the suitability of NHCs as spectator ligands in another Nobel-Prize-winning class of catalytic reactions: cross-coupling. In these carbon–carbon and carbon–heteroatom bond-forming processes, which are mediated by two-electron M^n/M^{n+2} redox cycles, the electronic and steric properties of NHCs lead to enhancements at various steps of the catalytic cycle (Fig. 5a). The strong σ-donating carbene leads to an electron-rich catalytically active metal centre that is better activated towards oxidative addition into carbon−halogen or pseudohalogen bonds of the substrates. This feature is particularly important for the coupling of challenging aryl chloride substrates, which possess strong carbon−chlorine bonds resistant to oxidative addition. The large steric influence of NHCs can also result in a more favourable reductive elimination step. Furthermore, both steric and electronic factors play a part in stabilizing the coordinatively unsaturated low-oxidation-state M^n active catalyst and reducing decomposition to heterogeneous material such as palladium black, which is often a major problem in these processes.

Since the initial report by Herrmann and co-workers[40], a huge array of different NHC–metal complexes have been prepared and employed as highly active and robust catalysts in a multitude of different coupling transformations. Reflecting the general trend in cross-coupling reactions, the majority of studies have focused on palladium-catalysed processes[48–51] although NHC complexes of many other transition metals including nickel[58] and iron[59] have also demonstrated remarkable efficiency. Imidazol-2-ylidene and imidazolin-2-ylidene carbenes remain the most studied classes, with sterically bulky derivatives such as IPr (**1d**) and SIPr (**2b**) being particularly widely applied. In many cases, the catalytically active NHC–metal complexes are prepared *in situ* by employing a suitable metal pre-cursor in the presence of an azolium salt, although many pre-formed, bench-stable NHC–metal pre-catalysts have been developed. Pioneering studies on such complexes were conducted by Nolan (reviewed in ref. 60), while the series of palladium(II) Pd–PEPPSI–NHC complexes **17** (Fig. 5b) introduced by Organ and co-workers have received widespread attention[61]. Dissociation of the stabilizing 3-chloropyridine ligand upon *in situ* reduction to palladium(0) delivers NHC-stabilized species, which are highly active catalysts for a variety of coupling processes, including the Negishi, Suzuki–Miyaura, Buchwald–Hartwig and Kumada reactions.

The operational simplicity and versatility of NHC synthesis is also a major factor in their success as spectator ligands in homogeneous catalysis. As well as enabling facile tuning of the electronic and steric properties of the active catalyst, this feature allows for the easy development of water-soluble[62] or immobilized derivatives[63] that increase the attractiveness of the catalysis reactions. Another growing area of research involves the use of chiral NHC–transition-metal complexes in asymmetric catalysis[64]. To generate a rigid, well defined chiral environment at the catalytically active metal centre, strategies must be employed that overcome the inherent anisotropic steric properties of NHCs by restricting or nullifying the effect of rotation around the metal–carbene and/or carbon–nitrogen bonds. Successful approaches include the use of C_2-symmetric NHCs, where rotation of the nitrogen substituents is impeded either by tethering or by steric repulsion by the backbone substituents. Alternatively, NHC ligands bearing an additional coordinating moiety can chelate the metal centre and fix the chiral environment. Notable applications of these concepts can be found in the field of asymmetric homogeneous hydrogenation. Highly efficient iridium olefin hydrogenation catalysts bearing chelating NHCs have been developed by the groups of Burgess[65] and, recently, Pfaltz[66], while a ruthenium complex stabilized by two chiral imidazolin-2-ylidene ligands reported by our group has emerged as an excellent catalyst for the asymmetric hydrogenation of various heteroarenes[67].

The strong σ-donating properties attractive for binding to individual metal centres in complexes are similarly beneficial to coordination to heterogeneous metallic materials. In contrast to the huge number of studies and widespread use of monometallic NHC species, however, the effect of NHC coordination on metallic surfaces has been markedly less well explored. An especially intriguing application in this area concerns their potential to stabilize transition-metal nanoparticles, which find multiple uses as heterogeneous catalysts. In 2010, our group described the combination of asymmetric and heterogeneous NHC–transition-metal catalysis by employing a chiral imidazolin-2-ylidene as a modifier for Fe_3O_4-supported palladium nanoparticles[68]. A pioneering study by Chaudret and co-workers in 2011 focused on the stabilization of ruthenium nanoparticles in the absence of a support, using the imidazol-2-ylidene ligands IPr (**1d**) and I*t*Bu (**1e**)[69]. Decomposition of a ruthenium(0) precursor in the presence of 0.2 equiv. (only for **1d**) or 0.5 equiv. of the free carbenes and H_2 gas (at 3 bar) in pentane delivered non-agglomerated, monodisperse colloids **18** with well defined sizes depending on the NHC concentration (about 1.5–1.7 nm, Fig. 6a). Attempted displacement of **1d** or **1e** by diphosphine or thiol ligands was unsuccessful, signifying the strength of NHC–metal binding and demonstrating the potential of these ligands as stabilizers for nanoparticles. ^{13}C nuclear-magnetic-resonance studies not only confirmed that coordination to the ruthenium surface occurs through the carbene carbon (C^2), but also revealed that the NHCs favoured the edge sites in a hexagonal-close-packed nanoparticle structure. The presence of free face sites available for catalysis was then demonstrated by the competence of colloids **18a** and **18b**

©2014 Macmillan Publishers Limited. All rights reserved

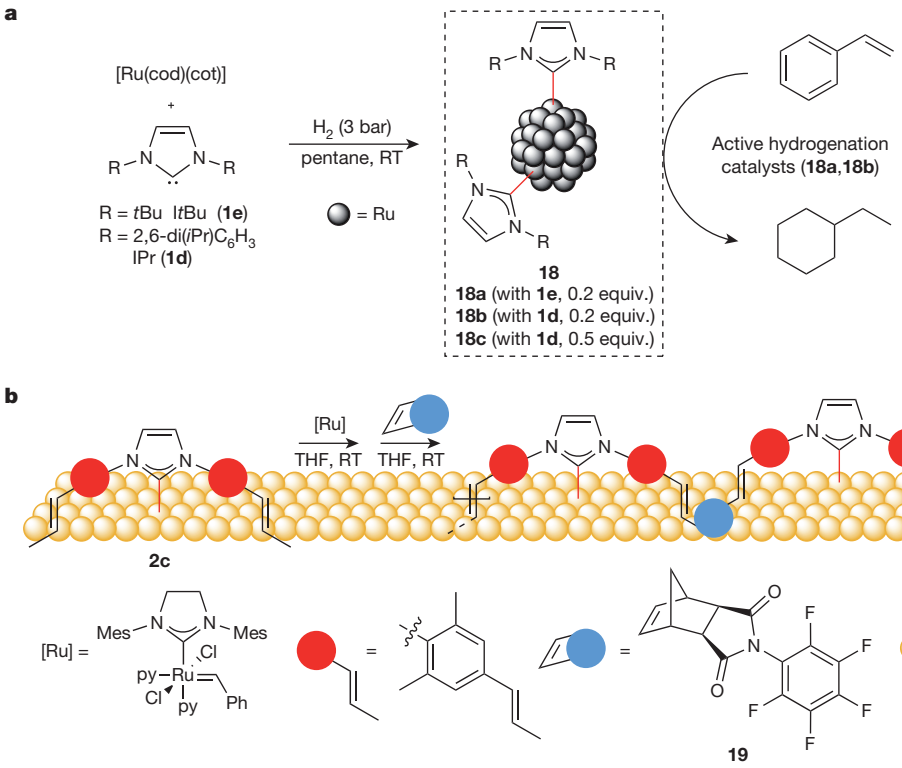


Figure 6 | NHC coordination to nanoparticles and metal surfaces. **a**, Preparation and catalytic reactivity of NHC-stabilized Ru-nanoparticles prepared by Chaudret and co-workers[69]. NHCs are capable of binding to heterogeneous metallic species in the same fashion as to monomeric complexes leading to nanoparticles that benefit from the high degree of stabilization afforded by the strength of the metal-NHC coordination. **b**, Modification of NHC ligands coordinated to a gold surface reported by Johnson and co-workers[70]. The strength of the binding between the NHC ligands and the metallic surface allowed for cross-metathesis to be performed directly on the surface-bound ligands, leading to functionalized materials. cod, 1,5-cyclooctadiene, cot, 1,3,5,7-cyclooctatetraene ; RT, room temperature; py, pyridine.

as catalysts in a representative arene hydrogenation reaction of styrene. Nanoparticles **18c**, prepared under a higher concentration of the less sterically demanding ligand IPr (**1d**), were shown also to possess NHCs bound to face sites and, as a result, displayed lower catalytic activity.

The potential of NHC binding to heterogeneous transition metal surfaces was further demonstrated in a recent 2013 study by Johnson and co-workers[70]. Well characterized monolayers of imidazolin-2-ylidene NHCs such as **2c** were prepared on a gold surface coated onto a silicon wafer. Subsequent immobilization of an NHC–ruthenium(II) complex via cross-metathesis with the alkene substituent in **2c** delivered a functional material susceptible to further reaction with the ROMP substrate **19** to form a surface functionalized with a polymer brush (a series of polymers attached at one end to a surface) (Fig. 6b). The remarkable stability of the ligand–gold binding throughout this sequence implies that suitably substituted NHCs could become useful anchor compounds that allow facile modification of the material properties via further functionalization.

Coordination of NHCs to *p*-block elements

Although the majority of reports have focused on their coordination to metals, NHCs also form adducts with a wide-range of non- or semi-metallic species (Fig. 7)[30,71]. As for NHC–metal complexes, σ-donation from the carbene into a vacant σ-accepting orbital of the low-valent *p*-block element is the dominant feature of these interactions. The strength of this dative coordination results in highly stable, non-labile complexes, which can exhibit dramatically different properties and reactivity to other related adducts. For example, in NHC–borane complexes **20** (Fig. 8a), the carbene–boron interaction can be considered as a covalent bond and the classical hydroboration reactivity of ether- or amine-borane adducts resulting from *in situ* generation of the uncoordinated parent borane is not observed[72]. Instead, these compounds can act as bases or nucleophiles, with hydroboration of ketones occurring only in the presence of an additional Lewis or Brønsted acid. The combination of sterically encumbered NHCs and bulky, electrophilic boranes such as $B(C_6F_5)_3$ does not lead to adducts but rather to frustrated Lewis pairs, which are capable of splitting hydrogen and other small molecules[73].

The strong coordination ability of NHCs has allowed the unprecedented stabilization of *p*-block elements in the zero oxidation state. The role of NHCs in these adducts has been compared to that of ligands in low-oxidation-state transition-metal complexes with σ-donation into vacant orbitals leading to enhanced stability[74]. The steric and electronic properties of NHCs have been found to have a great influence on the nature of adducts formed upon activation of white phosphorus with species featuring P_4, P_8 and P_{12} clusters all having been obtained using different carbenes. By contrast, reduction of an imidazol-2-ylidene–PCl_3 adduct by Robinson and co-workers led to the diatomic P_2 species **21**, which features phosphorus in a form reminiscent of that of nitrogen in N_2 (ref. 75). The same group also reported the first complex featuring silicon in the zero oxidation state (**22**) upon reduction of a silicon(IV) precursor (Fig. 8b)[76]. Similar adducts involving non-metals in the elemental oxidation state have been prepared for other *p*-block elements such as arsenic and, notably, carbon, with compound **23** being best described as a 'bent allene' with an

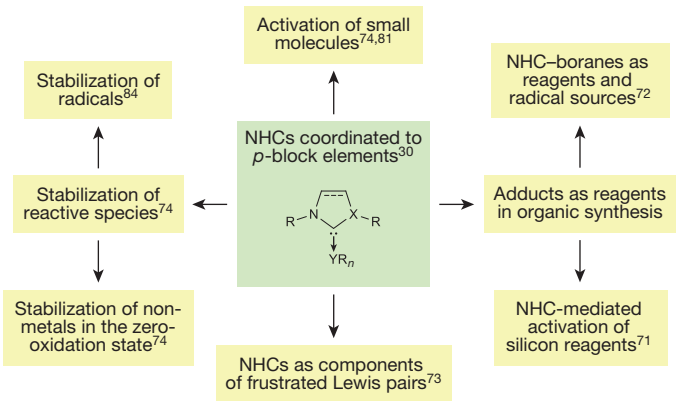


Figure 7 | Major applications of NHCs coordinated to *p*-block elements. NHCs coordinate strongly to a multitude of different *p*-block species, leading to adducts with a variety of different structures. NHCs have enabled the preparation and characterization of previously unknown species featuring *p*-block species in unconventional forms, such as in the zero-oxidation state or as radicals. NHCs may also activate small molecules either as themselves or as components of frustrated Lewis pairs.

©2014 Macmillan Publishers Limited. All rights reserved

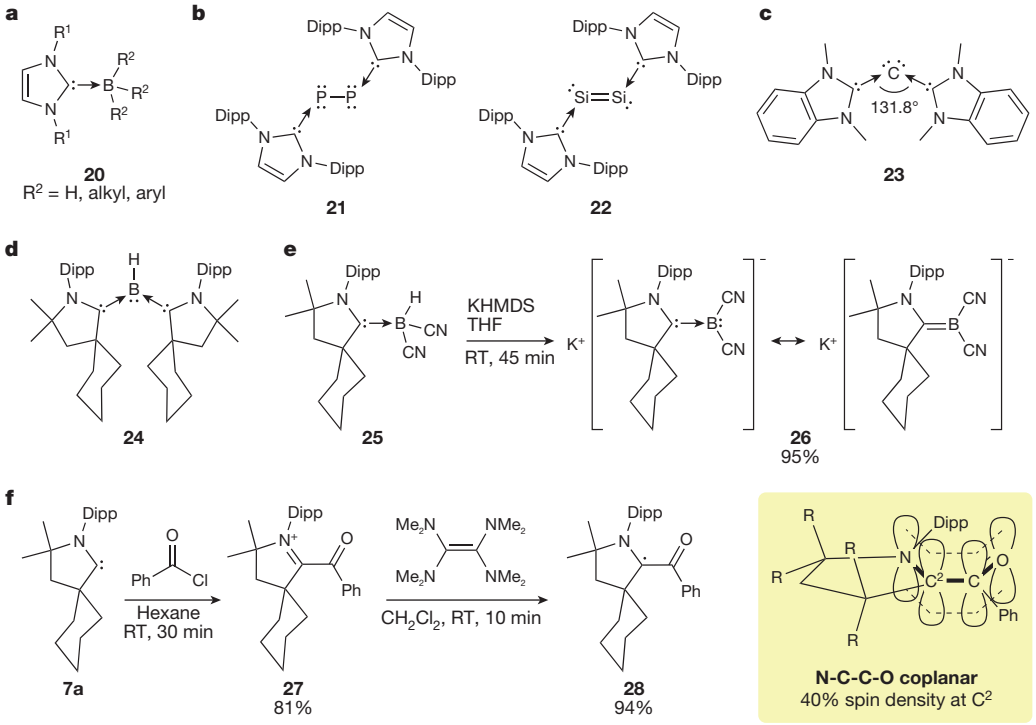


Figure 8 | Stabilization of *p*-block species by NHCs. **a**, Structure of NHC–borane complexes, **20**. These species find applications as reagents in organic synthesis and as co-initiators for polymerization. **b**, Structure of phosphorus(0) and silicon(0) species **21** and **22**, prepared by Robinson and co-workers[75,76]. **c**, Structure of 'bent allene' **23** reported by Bertrand and co-workers[77]. **d**, Structure of neutral tricoordinate boron species **24** (ref. 79). The boron atom in this adduct possesses a lone pair. **e**, Deprotonation of CAAC-coordinated borohydride **25**, recently reported by Bertrand and co-workers[80]. The π-electron-withdrawing abilities of the CAAC in this species allow for a formal inversion of the typical reactivity of borohydrides. **f**, Preparation of CAAC-stabilized organic radical **28** by Bertrand and co-workers[86]. The spin density in **28** is partially delocalized onto the CAAC with 40% being at the formerly carbene carbon C^2. Yields are also given (in per cent). Dipp, 2,6-(diisopropyl)phenyl; KHMDS, potassium hexamethyldisilazide; THF, tetrahydrofuran.

electronic structure featuring two lone pairs at the central carbon (Fig. 8c)[77]. Adducts between NHCs and fullerenes have also been prepared[78].

The stability and reactivity of NHC–*p*-block element adducts is also influenced by the π-properties of the NHC. In borane complexes **20**, the poor π-accepting ability of the carbene disfavours 1,2-migration of a boron substituent onto the adjacent electrophilic carbon. In other cases, however, the capacity to delocalize π-electron-density onto the NHC has helped enable the generation of previously elusive reactive species. Notable examples of this concept can also be found among NHC–boron adducts, where both borenium cations and boryl anions partially stabilized by π-interactions with imidazol-2-ylidene Lewis bases have been prepared[72]. Current research efforts in this area have focused on more recently developed classes of NHC, which offer improved characteristics for stabilizing reactive intermediates. Significant advances have been made by the Bertrand group using cyclic (alkyl)(amino)carbenes (CAACs, **7**). These NHCs, which feature only one nitrogen heteroatom in the ring, are considerably more π-accepting than imidazol-2-ylidenes and have enabled the preparation of unprecedented species such as the tri-coordinate neutral adduct **24**, which possesses a lone pair at boron (Fig. 8d)[79]. Moreover, CAAC **7a** in complex **25** was recently found to be sufficiently electron-withdrawing to invert the traditional polarity of boron–hydrogen bonds[80]. Treatment of this complex with potassium hexamethyldisilazide resulted in deprotonation at boron and delivered the isolable boryl anion **26** (Fig. 8e).

With filled σ-frontier and vacant π-frontier orbitals, the electronic structure of NHCs somewhat resembles that of transition metals and NHCs may exhibit similar reactivity to metals in the activation of small molecules[81]. π-Accepting classes of NHCs such as CAACs (**7**) and the carbonyl-containing DACs (**9**) are particularly well suited to these processes. As revealed in a series of reports by Bielawski and co-workers, these latter NHCs are capable of activating small molecules such as NH_3, while [2 + 1] cycloaddition reactivity analogous to that of classical transient carbenes has been observed with electron-neutral alkenes and alkynes[82,83].

Another interesting aspect of NHC coordination to *p*-block elements that is attracting an increasing amount of research attention is the stabilization of main-group radicals[84]. The π-accepting capability of NHCs once again has an important role in these systems, because delocalization of spin density across the adduct can result in greater stability. NHC-coordinated boryl radicals derived from complexes **20** were first described by Fensterbank, Lacôte, Malacria and Curran and co-workers in 2008, and have since found applications in organic radical chemistry and as co-initiators for radical polymerization[85]. Phosphorus, silicon and arsenic radicals have also been prepared with CAACs (**7**) proving especially adept at stabilizing these species. Very recently, a seminal report by Bertrand and co-workers described the preparation of an isolable, 'bottleable' carbon radical stabilized by CAAC, **7a**[86]. Treatment of the iminium salt **27** formed upon addition of the free NHC to benzoyl chloride with the one-electron reductant tetrakis(dimethylamino)ethylene led cleanly to the corresponding radical species **28** (Fig. 8f). As demonstrated by X-ray crystallographic analysis and calculations, the spin density in this adduct is partially delocalized onto the NHC with the N−C−C−O fragment being coplanar to maximize π-overlap. The remarkable stability of complex **28**, which could be stored for weeks at room temperature without decomposition, demonstrates the exceptional ability of NHCs, and specifically CAACs, to stabilize highly reactive intermediates and enable the structure and reactivity of previously inaccessible species to be studied directly.

NHCs as organocatalysts

Their propensity to coordinate to carbon-electrophiles has led to a third major class of applications, in which NHCs act as organocatalysts (Fig. 9)[87,88]. The majority of these processes are initiated by nucleophilic attack of the carbene onto carbonyl groups present in organic substrates.

©2014 Macmillan Publishers Limited. All rights reserved

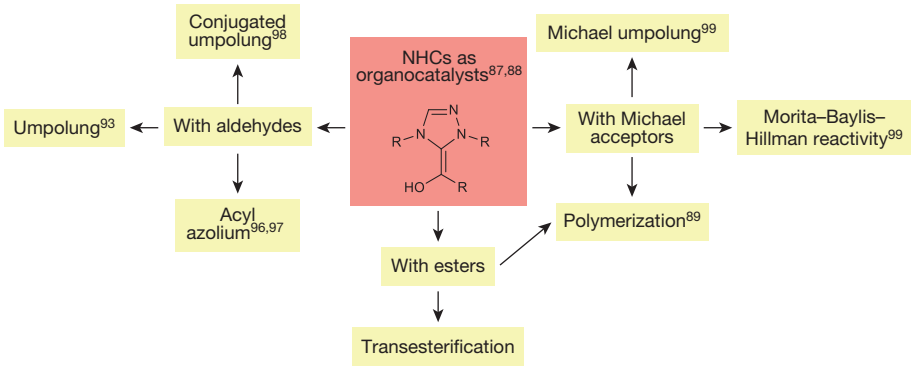


Figure 9 | Major applications of NHCs in organocatalysis. As organocatalysts, NHCs mediate a wide range of different organic transformations, with most processes involving an initial attack of the NHC onto a carbonyl group. Alongside transesterification and related transformations of esters, which are of particular relevance in polymer synthesis, the majority of NHC-catalysed reactions employ aldehydes as substrates. These processes involve an umpolung of the functional group with the carbonyl carbon acting as a transient nucleophile rather than an electrophile. Related transformations involving an umpolung at the β-position of α,β-unsaturated substrates are also known. These include 'conjugate umpolung' reactions of α,β-unsaturated aldehydes and processes involving direct attack of the NHC onto α,β-unsaturated esters (Michael umpolung). A further class of transformations involve azolium intermediates formed from aldehydes upon *in situ* oxidation or that have leaving groups in the α-position.

The electron-withdrawing nature of the cationic N-heterocyclic fragment generated upon nucleophilic attack has a key role in the subsequent reactivity of the adduct. In the case of esters, addition of the NHC to the carbonyl followed by release of the alkoxy group gives rise to an acyl azolium salt. This species is significantly more electrophilic than the original ester and can react with alcohols to afford transesterification products. Reactions of this type have found widespread use in step growth and ring-opening polymer synthesis, where NHCs offer an alternative to traditional organometallic catalysts and initiators[89]. Another role of NHCs in these processes results from their high Brønsted basicity with hydrogen bonding to the alcohol activating it towards nucleophilic attack.

The largest and most diverse array of NHC-organocatalysed reactions result from nucleophilic attack of NHCs on aldehydes. Although not well understood at the time, the first example of this kind of transformation dates back to 1943 when Ukai and co-workers reported the homodimerization of aldehydes to benzoins catalysed by a thiazolium salt[90]. As proposed by Breslow in 1958 (ref. 91), the mechanism of this process relies on the amphiphilic nature of an NHC active catalyst generated *in situ*. After initial nucleophilic attack of the NHC on the aldehyde, the formerly aldehydic proton in the resulting adduct is rendered acidic by the negative inductive effect of the cationic azolium group. Proton transfer then leads to the enamine-like 'Breslow intermediate' **29**, which is nucleophilic at carbon as a result of π-donation from the ring heteroatoms. The involvement of these species in NHC organocatalysis has been recently supported by the isolation and characterization of representative examples derived from SIPr (**2b**) by Berkessel and co-workers[92]. In the case of the benzoin condensation described above, nucleophilic attack of intermediate **29** onto another equivalent of the aldehyde followed by elimination of the NHC leads to product formation. During the course of the reaction, the innate reactivity of the aldehyde substrate is effectively inverted (that is, 'umpolung') with the normally electrophilic carbonyl carbon acting as a transient nucleophile. Transformations of this type are therefore examples of umpolung reactions and Breslow intermediates can be thought of as acyl anion equivalents[93].

The field of NHC-catalysed umpolung has grown rapidly over the last decades and a summary of the major reaction classes is shown in Fig. 10[87,88]. As for transformations catalysed by NHC–transition-metal complexes, there is generally no requirement to pre-form the free carbene in these processes. Instead, the active catalyst is typically generated *in situ* via deprotonation of the corresponding azolium salt precursor. The synthetic utility of many of these transformations has been greatly enhanced by the development of asymmetric variants using chiral NHCs. Catalysts based on the triazol-2-ylidene motif **5** have proved most effective in inducing enantioselectivity with systems featuring a chiral nitrogen substituent incorporated into a rigid polycyclic structure being widely employed. Using these unsymmetrical catalysts, one geometric isomer of the Breslow intermediate is favoured over the other while approach of the electrophile is directed to the least hindered enantiotopic face. Under conditions where the benzoin condensation is reversible, nucleophilic attack of the Breslow intermediate onto other electrophiles may also occur. In particular, hydroacylation reactions involving addition of aldehydes to activated alkenes such as Michael acceptors have been extensively studied (the Stetter reaction). Extending the range of suitable olefinic coupling partners in this process has been a focus of research in our group[94], with the successful application of comparatively electron-neutral styrene derivatives being recently realized using an electron-rich 2,6-dimethoxyphenyl N-substituted triazol-2-ylidene catalyst[95].

Alternative reactivity pathways of Breslow intermediates not involving a formal umpolung at the carbonyl carbon are also accessible[96]. For example, elimination of a leaving group situated in the α-position can generate acyl azolium salts of the same type that is observed upon NHC addition to esters. Similar species may also be formed from pre-oxidized substrates such as ketenes or upon direct *in situ* oxidation of the Breslow intermediate in the presence of an external oxidant. These intermediates may release the NHC fragment directly upon addition-elimination of a nucleophile or first react as enolate or enone equivalents with the azolium moiety as a bystander[97]. A particularly powerful and widely studied class of transformations concerns the reactivity of α,β-unsaturated aldehydes. The Breslow-type intermediates formed with these substrates possess an extended π-system and, consequently, nucleophilic attack may occur in a conjugate fashion to afford products resulting from an umpolung at the β-position (termed "conjugate umpolung")[98]. Sterically demanding NHC catalysts, which reduce competitive functionalization at the classical 'carbonyl' position, are often beneficial in these processes. Another class of reactions with α,β-unsaturated carbonyl compounds involve the initial conjugate addition of the NHC to the β-position rather than to the carbonyl group. The resulting adducts can then react to afford α- or β-functionalized products resulting from Morita–Baylis–Hillman-type chemistry or a formal umpolung at the β-position, respectively[99].

One especially exciting area of current research activity involves accessing the umpolung reactivity observed with aldehydes from other classes of substrates. An elegant demonstration of this concept was recently reported by Chi and co-workers in a variety of annulation reactions of saturated aliphatic esters **30** (ref. 100). A mechanistic rationale for these processes, showing the formation of the key diamino dienol Breslow-type intermediate **31**, is illustrated in Fig. 11a. After initial nucleophilic addition-elimination of the NHC to the ester group, tautomerization of the resulting acyl azolium salt gives rise to an enol species **32**. The electron-withdrawing nature of the azolium substituent in this compound renders the β-CH_2 protons relatively acidic and, in the presence of excess DBU base (where

©2014 Macmillan Publishers Limited. All rights reserved

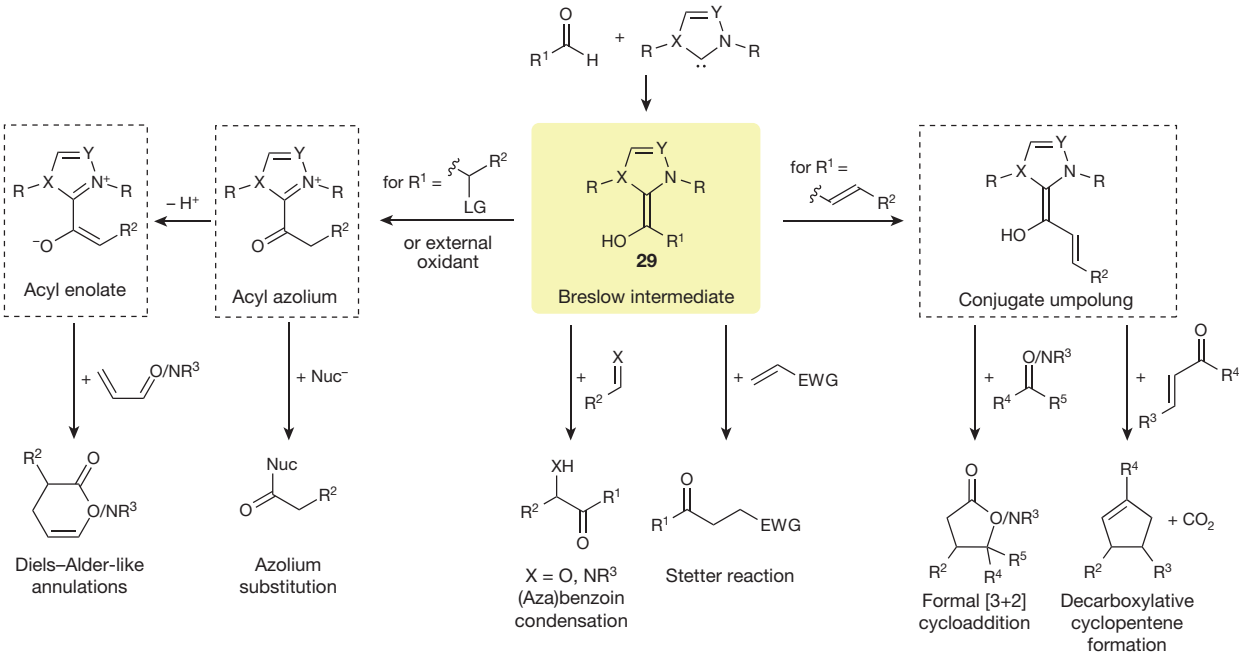


Figure 10 | Major NHC-catalysed reactions of aldehydes. The reaction between NHC organocatalysts and aldehydes leads to the enamine-like 'Breslow intermediate' **29**. This species is nucleophilic at the formerly carbonyl carbon and reacts to form products resulting from an umpolung of the aldehyde. The two most studied transformations of this type are the (aza)benzoin condensation where the Breslow intermediate attacks an aldehyde or imine and the Stetter reaction involving nucleophilic attack onto an electron-deficient alkene. With α,β-unsaturated aldehydes, the Breslow-type intermediate can instead act as a nucleophile at the β-position (termed 'conjugate umpolung') leading to annulated products. A third major reaction class involves acyl azolium intermediates. These may be formed directly upon elimination of Breslow-type intermediates derived from aldehydes bearing leaving groups at the α-position, or upon oxidation with an external oxidant. Subsequent trapping by a nucleophile leads to products of oxidation, although other reactivity pathways are also accessible where the azolium moiety acts as a spectator. EWG, electron-withdrawing group; LG, leaving group; Nuc, nucleophile.

DBU is 1,8-diazabicyclo[5.4.0]undec-7-ene, 1.5–2 equiv.), deprotonation can occur to afford the homoenolate equivalent **31**. The remarkable selectivity for this deprotonation step over more conventional α-functionalization pathways can also be partly explained by the high degree of conjugation present in the resulting intermediate with β-aryl-substituted substrates **30**. The Breslow-type species **31** is analogous to that obtained under conventional conditions with α,β-unsaturated aldehyde substrates and can react as a nucleophile via conjugate umpolung to afford β-functionalized products.

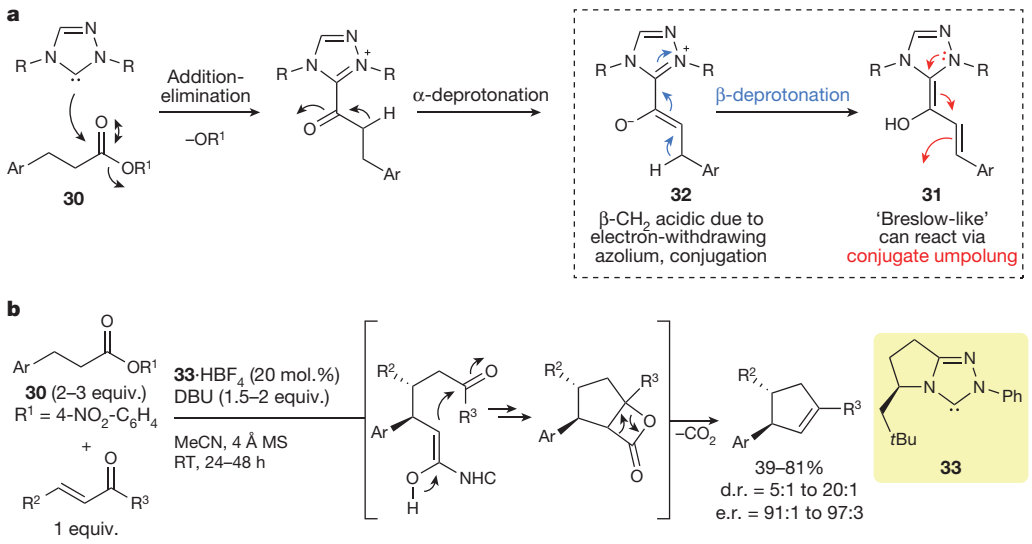


Figure 11 | NHC-catalysed β-functionalization of saturated esters. Recently reported by Chi and co-workers[100], this reaction involves accessing NHC-catalysed reactivity pathways conventionally observed with aldehydes from saturated esters. **a**, Mechanism for the formation of 'Breslow-like' diamino dienol intermediate **31** from saturated aliphatic esters **30**. A key deprotonation at the β-position of the acyl azolium intermediate formed upon addition-elimination to the ester (shown in blue) leads to the same kind of intermediate that is observed upon direct addition of NHCs to α,β-unsaturated aldehydes. **b**, NHC-catalysed asymmetric annulation of saturated aliphatic esters and enones. Nucleophilic attack of the Breslow-type intermediate **31** onto the enone followed by ring closing and extrusion of carbon dioxide leads to cyclopentenes. High levels of enantioselectivity were observed using the chiral triazolylidene catalyst **33**. DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene; e.r., enantiomeric ratio; MeCN, acetonitrile; HBF_4, tetrafluoroboric acid; d.r., diastereoisomeric ratio; MS, molecular sieves.

©2014 Macmillan Publishers Limited. All rights reserved

Despite the prevalence and widespread use of aliphatic esters as feedstocks in organic chemistry, alternative synthetic methods of selectively activating the β-position of these substrates are scarce. Moreover, using the chiral triazol-2-ylidene catalyst **33**, this methodology allows for high levels of enantiocontrol during the annulation reactions with five-membered carbo- and heterocycles being delivered in enantiomeric ratios of up to 97:3, using enone, trifluoroketone or hydrazone electrophiles (the reaction with enones is shown in Fig. 11b).

Outlook

The discovery and development of N-heterocyclic carbenes is undoubtedly one of the greatest success stories of recent chemistry research. In the 23 years since Arduengo and co-workers reported the first 'bottleable' NHC, seminal contributions from many different groups on the structure, coordination chemistry and reactivity of these compounds have led to a multitude of applications across many different fields. NHCs are now workhorses of organic and organometallic chemistry, rivalling phosphines as ancillary ligands in transition metal catalysis and offering new possibilities in main-group chemistry and organocatalysis.

However, as we hope is clear from the examples of recent research touched upon in this review, the meteoric rise of NHCs is far from complete. Alongside their other roles, NHCs are continuing to find many new applications across the chemical sciences. One area of promise is the field of heterogeneous catalysis, where the strength of NHC–metal binding can allow for enhanced stabilization of metallic colloids or surfaces, with the potential for modification of the catalyst properties via *in situ* functionalization of the ligand. The strength and stability of the metal–ligand bond, and the ability to readily fine-tune the properties of organometallic complexes through structural modification of the ligand, also explain the increasing use of NHCs in metallopharmaceuticals. Significant advances are also being made in fields where the use of NHCs is more established. In organocatalysis, recent pioneering research has focused on the development of new reactivity pathways which expand the range of suitable reaction partners beyond the traditional aldehydes.

A major driver of current ground-breaking research is the development of new NHCs with different properties and reactivities. In the field of homogeneous transition-metal catalysis, ruthenium catalysts featuring novel chelating imidazolin-2-ylidene ligands have facilitated *Z*-selective olefin metathesis, while recently developed chiral NHC ligands have shown promise in asymmetric hydrogenation reactions. New classes of NHCs such as DACs and CAACs have shown unprecedented reactivity in the activation of small molecules and in the stabilization of hitherto inaccessible non-metallic species. CAACs are particularly suited to such stabilization, as exemplified by the recent report of a CAAC-stabilized 'bottleable' organic radical. Given the tremendous strides that have been taken over the last two decades, and the high-quality research currently being conducted, the future of NHCs looks very exciting.

Received 10 January; accepted 9 April 2014.

1. Dumas, J. B. & Peligot, E. Mémoire sur l'esprit-de-bois et les divers composés éthéres qui en proviennent. *Ann. Chim. Phys.* **58,** 5–74 (1835).
2. Arduengo, A. J., III & Krafczyk, R. Auf der Suche nach stabilen Carbenen. *Chem. Unserer Zeit* **32,** 6–14 (1998).
3. Igau, A., Grutzmacher, H., Baceiredo, A. & Bertrand, G. Analogous α,α′-bis-carbenoid triply bonded species: synthesis of a stable λ^3-phosphinocarbene-λ^5-phosphaacetylene. *J. Am. Chem. Soc.* **110,** 6463–6466 (1988).
4. Arduengo, A. J., III, Harlow, R. L. & Kline, M. A stable crystalline carbene. *J. Am. Chem. Soc.* **113,** 361–363 (1991).
 This is the first report of a stable, isolable NHC.
5. Wanzlick, H.-W. & Schönherr, H.-J. Direct synthesis of a mercury salt-carbene complex. *Angew. Chem. Int. Edn Engl.* **7,** 141–142 (1968).
6. Öfele, K. 1,3-Dimethyl-4-imidazolinyliden-(2)-pentacarbonylchrom Ein Neuer Übergangsmetall-carben-komplex. *J. Organomet. Chem.* **12,** P42–P43 (1968).
7. Bourissou, D., Guerret, O., Gabbaï, F. P. & Bertrand, G. Stable carbenes. *Chem. Rev.* **100,** 39–92 (2000).
8. de Frémont, P., Marion, N. & Nolan, S. P. Carbenes: synthesis, properties, and organometallic chemistry. *Coord. Chem. Rev.* **253,** 862–892 (2009).
9. Herrmann, W. A. & Köcher, C. N-heterocyclic carbenes. *Angew. Chem. Int. Edn Engl.* **36,** 2162–2187 (1997).
10. Runyon, J. W. *et al.* Carbene-based Lewis pairs for hydrogen activation. *Aust. J. Chem.* **64,** 1165–1172 (2011).
11. Heinemann, C., Müller, T., Apeloig, Y. & Schwarz, H. On the question of stability, conjugation, and "aromaticity" in imidazol-2-ylidenes and their silicon analogs. *J. Am. Chem. Soc.* **118,** 2023–2038 (1996).
12. Arduengo, A. J., III, Rasika Dias, H. V., Harlow, R. L. & Kline, M. Electronic stabilization of nucleophilic carbenes. *J. Am. Chem. Soc.* **114,** 5530–5534 (1992).
13. Arduengo, A. J., III, Goerlich, J. R. & Marshall, W. J. A stable diaminocarbene. *J. Am. Chem. Soc.* **117,** 11027–11028 (1995).
14. Melaimi, M., Soleilhavoup, M. & Bertrand, G. Stable cyclic carbenes and related species beyond diaminocarbenes. *Angew. Chem. Int. Edn* **49,** 8810–8849 (2010).
15. Lavallo, V., Canac, Y., Präsang, C., Donnadieu, B. & Bertrand, G. Stable cyclic (alkyl)(amino)carbenes as rigid or flexible, bulky, electron-rich ligands for transition-metal catalysts: a quaternary carbon atom makes the difference. *Angew. Chem. Int. Edn* **44,** 5705–5709 (2005).
 This is the first report of a cyclic (alkyl)(amino)carbene (CAAC).
16. Aldeco-Perez, E. *et al.* Isolation of a C5-deprotonated imidazolium, a crystalline "abnormal" N-heterocyclic carbene. *Science* **326,** 556–559 (2009).
17. Schuster, O. & Yang, L. Raubenheimer, H. G. & Albrecht, M. Beyond conventional *N*-heterocyclic carbenes: abnormal, remote and other classes of NHC ligands with reduced heteroatom stabilization. *Chem. Rev.* **109,** 3445–3478 (2009).
18. Benhamou, L., Chardon, E., Lavigne, G., Bellemin-Laponnaz, S. & César, V. Synthetic routes to N-heterocyclic carbene precursors. *Chem. Rev.* **111,** 2705–2733 (2011).
19. Dröge, T. & Glorius, F. The measure of all rings – N-heterocyclic carbenes. *Angew. Chem. Int. Edn* **49,** 6940–6952 (2010).
20. Nelson, D. J. & Nolan, S. P. Quantifying and understanding the electronic properties of *N*-heterocyclic carbenes. *Chem. Soc. Rev.* **42,** 6723–6753 (2013).
 This recent review gives a comprehensive overview of the electronic properties of NHCs and includes data for about 300 compounds.
21. Hillier, A. C. *et al.* A combined experimental and theoretical study examining the binding of *N*-heterocyclic carbenes (NHC) to the Cp*RuCl (Cp* = η^5-C_5Me_5) moiety: insight into stereoelectronic differences between unsaturated and saturated NHC ligands. *Organometallics* **22,** 4322–4326 (2003).
22. Tolman, C. A. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chem. Rev.* **77,** 313–348 (1977).
23. Cardin, D. J., Çetinkaya, B. & Lappert, M. F. Transition metal-carbene complexes. *Chem. Rev.* **72,** 545–574 (1972).
24. Díez-González, S. & Nolan, S. P. Stereoelectronic parameters associated with N-heterocyclic carbene (NHC) ligands: a quest for understanding. *Coord. Chem. Rev.* **251,** 874–883 (2007).
25. Jacobsen, H., Correa, A., Poater, A., Costabile, C. & Cavallo, L. Understanding the M-(NHC) (NHC = N-heterocyclic carbene) bond. *Coord. Chem. Rev.* **253,** 687–703 (2009).
26. Nemcsok, D., Wichmann, K. & Frenking, G. The significance of π interactions in Group 11 complexes with N-heterocyclic carbenes. *Organometallics* **23,** 3640–3646 (2004).
27. Crabtree, R. H. NHC ligands versus cyclopentadienyls and phosphines as spectator ligands in organometallic chemistry. *J. Organomet. Chem.* **690,** 5451–5457 (2005).
28. Crudden, C. M. & Allen, D. P. Stability and reactivity of *N*-heterocyclic carbene complexes. *Coord. Chem. Rev.* **248,** 2247–2273 (2004).
29. Hahn, F. E. & Jahnke, M. C. Heterocyclic carbenes: synthesis and coordination chemistry. *Angew. Chem. Int. Ed.* **47,** 3122–3172 (2008).
 This is an excellent comprehensive review summarising the synthesis and coordination chemistry of a range of NHCs.
30. Kuhn, N. & Al-Sheikh, A. 2,3-Dihydroimidazol-2-ylidenes and their main group element chemistry. *Coord. Chem. Rev.* **249,** 829–857 (2005).
31. Arnold, P. L. & Casely, I. J. F-block N-heterocyclic carbene complexes. *Chem. Rev.* **109,** 3599–3611 (2009).
32. Poyatos, M., Mata, J. A. & Peris, E. Complexes with poly(N-heterocyclic carbene) ligands: structural features and catalytic applications. *Chem. Rev.* **109,** 3677–3707 (2009).
33. Mercs, L. & Albrecht, M. Beyond catalysis: *N*-heterocyclic carbene complexes as components for medicinal, luminescent, and functional materials applications. *Chem. Soc. Rev.* **39,** 1903–1912 (2010).
34. Oisaki, K., Li, Q., Furukawa, H., Czaja, A. U. & Yaghi, O. M. A. Metal-organic framework with covalently bound organometallic complexes. *J. Am. Chem. Soc.* **132,** 9262–9264 (2010).
35. Lee, K. M., Lee, C. K. & Lin, I. J. B. A facile synthesis of unusual liquid-crystalline gold(I) dicarbene compounds. *Angew. Chem. Int. Edn Engl.* **36,** 1850–1852 (1997).
36. Boydston, A. J., Williams, K. A. & Bielawski, C. W. A modular approach to main-chain organometallic polymers. *J. Am. Chem. Soc.* **127,** 12496–12497 (2005).
37. Visbal, R. & Concepción Gimeno, M. N-heterocyclic carbene metal complexes: photoluminescence and applications. *Chem. Soc. Rev.* **43,** 3551–3574 (2014).
38. Hindi, K. M., Panzner, M. J., Tessier, C. A., Cannon, C. L. & Youngs, W. J. The medicinal applications of imidazolium carbene–metal complexes. *Chem. Rev.* **109,** 3859–3884 (2009).
39. Hickey, J. L. *et al.* Mitochondria-targeted chemotherapeutics: the rational design of gold(I) *N*-heterocyclic carbene complexes that are selectively toxic to cancer cells and target protein selenols in preference to thiols. *J. Am. Chem. Soc.* **130,** 12570–12571 (2008).
40. Herrmann, W. A., Elison, M., Fischer, J., Köcher, C. & Artus, G. R. J. Metal complexes of N-heterocyclic carbenes—a new structural principle for catalysts in homogeneous catalysis. *Angew. Chem. Int. Edn Engl.* **34,** 2371–2374 (1995).
 This is the first report to apply an NHC as a ligand in transition-metal catalysis.
41. Díez-González, S., Marion, N. & Nolan, S. P. N-heterocyclic carbenes in late transition metal catalysis. *Chem. Rev.* **109,** 3612–3676 (2009).

©2014 Macmillan Publishers Limited. All rights reserved

This comprehensive review on NHCs as ligands in transition-metal catalysis forms part of a special issue of *Chemical Reviews* on carbenes.
42. Herrmann, W. A. N-heterocyclic carbenes: a new concept in organometallic catalysis. *Angew. Chem. Int. Ed.* **41,** 1290–1309 (2002).
43. Glorius, F. *N-Heterocyclic Carbenes in Transition Metal Catalysis* (Springer, 2007).
44. Nolan, S. P. *N-Heterocyclic Carbenes in Synthesis* (Wiley, 2006).
45. Normand, A. T. & Cavell, K. J. Donor-functionalised N-heterocyclic carbene complexes of Group 9 and Group 10 metals in catalysis: trends and directions. *Eur. J. Inorg. Chem.* 2781–2800 (2008).
46. Marion, N. & Nolan, S. P. N-heterocyclic carbenes in gold catalysis. *Chem. Soc. Rev.* **37,** 1776–1782 (2008).
47. Marciniec, B. Hydrosilylation. In *Advances in Silicon Science* (ed. Marciniec, B.) Vol. 1 3–51 (Springer, 2009).
48. Kantchev, E. A. B., O'Brien, C. J. & Organ, M. G. Palladium complexes of N-heterocyclic carbenes as catalysts for cross-coupling reactions—a synthetic chemists perspective. *Angew. Chem. Int. Edn* **46,** 2768–2813 (2007).
49. Fortman, G. C. & Nolan, S. P. *N*-heterocyclic carbene (NHC) ligands and palladium in homogeneous cross-coupling catalysis: a perfect union. *Chem. Soc. Rev.* **40,** 5151–5169 (2011).
50. Würtz, S. & Glorius, F. Surveying sterically demanding N-heterocyclic carbene ligands with restricted flexibility for palladium-catalyzed cross-coupling reactions. *Acc. Chem. Res.* **41,** 1523–1533 (2008).
51. Valente, C. *et al.* The development of bulky palladium NHC complexes for the most-challenging cross-coupling reactions. *Angew. Chem. Int. Edn* **51,** 3314–3332 (2012).
52. Vougioukalakis, G. C. & Grubbs, R. H. Ruthenium-based heterocyclic carbene-coordinated olefin metathesis catalysts. *Chem. Rev.* **110,** 1746–1787 (2010).
53. Samojłowicz, C., Bieniek, M. & Grela, K. Ruthenium-based olefin metathesis catalysts bearing *N*-heterocyclic carbene ligands. *Chem. Rev.* **109,** 3708–3742 (2009).
54. Sanford, M. S., Love, J. A. & Grubbs, R. H. Mechanism and activity of ruthenium olefin metathesis catalysts. *J. Am. Chem. Soc.* **123,** 6543–6554 (2001).
55. Endo, K. & Grubbs, R. H. Chelated ruthenium catalysts for *Z*-selective olefin metathesis. *J. Am. Chem. Soc.* **133,** 8525–8527 (2011).
56. Keitz, B. K., Endo, K., Patel, P. R., Herbert, M. B. & Grubbs, R. H. Improved ruthenium catalysts for *Z*-selective olefin metathesis. *J. Am. Chem. Soc.* **134,** 693–699 (2012).
57. Hartung, J. & Grubbs, R. H. Highly *Z*-selective and enantioselective ring-opening/cross-metathesis catalyzed by a resolved stereogenic-at-Ru complex. *J. Am. Chem. Soc.* **135,** 10183–10185 (2013).
58. Gu, S., Ni, P. & Chen, W. *N*-heterocyclic carbenes: versatile reagents for nickel-catalyzed coupling reactions. *Chin. J. Catal.* **31,** 875–886 (2010).
59. Bézier, D., Sortais, J.-B. & Darcel, C. N-heterocyclic carbene ligands and iron: an effective association for catalysis. *Adv. Synth. Catal.* **355,** 19–33 (2013).
60. Marion, N. & Nolan, S. P. Well-defined N-heterocyclic carbenes–palladium(II) precatalysts for cross-coupling reactions. *Acc. Chem. Res.* **41,** 1440–1449 (2008).
61. O'Brien, C. J. *et al.* Easily prepared air- and moisture-stable Pd-NHC (NHC = N-heterocyclic carbene) complexes: a reliable, user-friendly, highly active palladium precatalyst for the Suzuki-Miyaura reaction. *Chem. Eur. J.* **12,** 4743–4748 (2006).
62. Schaper, L.-A., Hock, S. J., Herrmann, W. A. & Kühn, F. E. Synthesis and application of water-soluble NHC transition-metal complexes. *Angew. Chem. Int. Edn* **52,** 270–289 (2013).
63. Ranganath, K. V. S., Onitsuka, S., Kumar, A. K. & Inanaga, J. Recent progress of *N*-heterocyclic carbenes in heterogeneous catalysis. *Catal. Sci. Technol.* **3,** 2161–2181 (2013).
64. Wang, F., Liu, L.-J., Wang, W., Li, S. & Shi, M. Chiral NHC-metal-based asymmetric catalysis. *Coord. Chem. Rev.* **256,** 804–853 (2012).
65. Powell, M. T., Hou, D.-R., Perry, M. C., Cui, X. & Burgess, K. Chiral imidazolylidine ligands for asymmetric hydrogenation of aryl alkenes. *J. Am. Chem. Soc.* **123,** 8878–8879 (2001).
66. Schumacher, A., Bernasconi, M. & Pfaltz, A. Chiral N-heterocyclic carbene/pyridine ligands for the iridium-catalyzed asymmetric hydrogenation of olefins. *Angew. Chem. Int. Edn* **52,** 7422–7425 (2013).
67. Ortega, N., Urban, S., Beiring, B. & Glorius, F. Ruthenium NHC catalyzed highly asymmetric hydrogenation of benzofurans. *Angew. Chem. Int. Edn* **51,** 1710–1713 (2012).
68. Ranganath, K. V. S., Kloesges, J., Schäfer, A. H. & Glorius, F. Asymmetric nanocatalysis: N-heterocyclic carbenes as chiral modifiers of Fe_3O_4/Pd nanoparticles. *Angew. Chem. Int. Edn* **49,** 7786–7789 (2010).
This paper reports the first asymmetric heterogeneous catalysis reaction employing NHCs as chiral modifiers.
69. Lara, P. *et al.* Ruthenium nanoparticles stabilized by N-heterocyclic carbenes: ligand location and influence on reactivity. *Angew. Chem. Int. Edn* **50,** 12080–12084 (2011).
70. Zhukhovitskiy, A. V., Mavros, M. G., Van Voorhis, T. & Johnson, J. A. Addressable carbene anchors for gold surfaces. *J. Am. Chem. Soc.* **135,** 7418–7421 (2013).
71. Fuchter, M. J. N-heterocyclic carbene mediated activation of tetravalent silicon compounds: a critical evaluation. *Chem. Eur. J.* **16,** 12286–12294 (2010).
72. Curran, D. P. *et al.* Synthesis and reactions of N-heterocyclic carbene boranes. *Angew. Chem. Int. Edn* **50,** 10294–10317 (2011).
This review provides an excellent summary of the synthesis and reactivity of NHC–borane adducts.
73. Kolychev, E. L., Theuergarten, E. & Tamm, M. N-heterocyclic carbenes in FLP chemistry. *Top. Curr. Chem.* **334,** 121–155 (2013).
74. Martin, D., Soleilhavoup, M. & Bertrand, G. Stable singlet carbenes as mimics for transition metal centers. *Chem. Sci.* **2,** 389–399 (2011).
75. Wang, Y. *et al.* Carbene-stabilized diphosphorus. *J. Am. Chem. Soc.* **130,** 14970–14971 (2008).
76. Wang, Y. *et al.* A stable silicon(0) compound with a Si=Si double bond. *Science* **321,** 1069–1071 (2008).
77. Dyker, C. A., Lavallo, V., Donnadieu, B. & Bertrand, G. Synthesis of an extremely bent acyclic allene (a ''carbodicarbene''): a strong donor ligand. *Angew. Chem. Int. Edn* **47,** 3206–3209 (2008).
78. Li, H. *et al.* Fullerene-carbene Lewis acid-base adducts. *J. Am. Chem. Soc.* **133,** 12410–12413 (2011).
79. Kinjo, R., Donnadieu, B., Celik, M. A., Frenking, G. & Bertrand, G. Synthesis and characterization of a neutral tricoordinate organoboron isoelectronic with amines. *Science* **333,** 610–613 (2011).
80. Ruiz, D. A., Ung, G., Melaimi, M. & Bertrand, G. Deprotonation of a borohydride: synthesis of a carbene-stabilized boryl anion. *Angew. Chem. Int. Edn* **52,** 7590–7592 (2013).
81. Delaude, L. Betaine adducts of N-heterocyclic carbenes: synthesis, properties, and reactivity. *Eur. J. Inorg. Chem.* 1681–1699 (2009).
82. Moerdyk, J. P. & Bielawski, C. W. Diamidocarbenes as versatile and reversible [2+1] cycloaddition reagents. *Nature Chem.* **4,** 275–280 (2012).
83. Moerdyk, J. P. & Bielawski, C. W. Alkyne and reversible nitrile activation: *N,N′*-diamidocarbene-facilitated synthesis of cyclopropenes, cyclopropenones, and azirines. *J. Am. Chem. Soc.* **134,** 6116–6119 (2012).
84. Martin, C. D., Soleilhavoup, M. & Bertrand, G. Carbene-stabilized main group radicals and radical ions. *Chem. Sci.* **4,** 3020–3030 (2013).
85. Ueng, S.-H. *et al.* Complexes of borane and N-heterocyclic carbenes: a new class of radical hydrogen atom donor. *J. Am. Chem. Soc.* **130,** 10082–10083 (2008).
86. Mahoney, J. K., Martin, D., Moore, C. E., Rheingold, A. L. & Bertrand, G. Bottleable (amino)(carboxy) radicals derived from cyclic (alkyl)(amino) carbenes. *J. Am. Chem. Soc.* **135,** 18766–18769 (2013).
87. Enders, D., Niemeier, O. & Henseler, A. Organocatalysis by N-heterocyclic carbenes. *Chem. Rev.* **107,** 5606–5655 (2007).
This review gives a broad overview of the applications of NHCs as organocatalysts.
88. Chiang, P.-C. & Bode, J. W. in *N-Heterocyclic Carbenes: From Laboratory Curiosities to Efficient Synthetic Tools* (ed. Díez-González, S.) 399–435 (Royal Society of Chemistry, 2011).
89. Fèvre, M., Pinaud, J., Gnanou, Y., Vignolle, J. & Taton, D. *N*-heterocyclic carbenes (NHCs) as organocatalysts and structural components in metal-free polymer synthesis. *Chem. Soc. Rev.* **42,** 2142–2172 (2013).
90. Ukai, T., Tanaka, R., Dokawa, T. A new catalyst for acyloin condensation. [in Japanese] *J. Pharm. Soc. Jpn* **63,** 296–300 (1943).
91. Breslow, R. On the mechanism of thiamine action. IV. Evidence from studies on model systems. *J. Am. Chem. Soc.* **80,** 3719–3726 (1958).
92. Berkessel, A. *et al.* Umpolung by N-heterocyclic carbenes: generation and reactivity of the elusive 2,2-diamino enols (Breslow intermediates). *Angew. Chem. Int. Edn* **51,** 12370–12374 (2012).
This paper provides experimental evidence for the involvement of Breslow intermediates in NHC organocatalysis.
93. Bugaut, X. & Glorius, F. Organocatalytic umpolung: *N*-heterocyclic carbenes and beyond. *Chem. Soc. Rev.* **41,** 3511–3522 (2012).
94. Biju, A. T., Kuhl, N. & Glorius, F. Extending NHC-catalysis: coupling aldehydes with unconventional reaction partners. *Acc. Chem. Res.* **44,** 1182–1195 (2011).
95. Schedler, M., Wang, D.-S. & Glorius, F. NHC-catalyzed hydroacylation of styrenes. *Angew. Chem. Int. Edn* **52,** 2585–2589 (2013).
96. Ryan, S. J., Candish, L. & Lupton, D. W. Acyl anion free *N*-heterocyclic carbene organocatalysis. *Chem. Soc. Rev.* **42,** 4906–4917 (2013).
97. Vora, H. U., Wheeler, P. & Rovis, T. Exploiting acyl and enol azolium intermediates via N-heterocyclic carbene-catalyzed reactions of α-reducible aldehydes. *Adv. Synth. Catal.* **354,** 1617–1639 (2012).
98. Nair, V. *et al.* Employing homoenolates generated by NHC catalysis in carbon-carbon bond-forming reactions: state of the art. *Chem. Soc. Rev.* **40,** 5336–5346 (2011).
99. Chen, X.-Y. & Ye, S. N-heterocyclic carbene-catalyzed reactions of C–C unsaturated bonds. *Org. Biomol. Chem.* **11,** 7991–7998 (2013).
100. Fu, Z., Xu, J., Zhu, T., Leong, W. W. Y. & Chi, Y. R. β-Carbon activation of saturated carboxylic esters through *N*-heterocyclic carbene organocatalysis. *Nature Chem.* **5,** 835–839 (2013).

Acknowledgements We thank the European Research Council under the European Community's Seventh Framework Program (FP7 2007-2013)/ERC grant agreement number 25936, the Deutsche Forschungsgemeinschaft (Leibniz Award and SFB 858), the Alexander von Humboldt Foundation (to M.N.H.) and the Fonds der Chemischen Industrie (to M.S.) for financial support.

Author Contributions All authors worked together to outline the content of the review and define its scope. The text was primarily written by M.N.H. and F.G. with contributions from all authors. The figures were prepared by M.N.H., C.R. and M.S. Editing of the manuscript, figures and references was done by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.G. (glorius@uni-muenster.de).

©2014 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature13457

A zero-knowledge protocol for nuclear warhead verification

Alexander Glaser[1], Boaz Barak[2] & Robert J. Goldston[3]

The verification of nuclear warheads for arms control involves a paradox: international inspectors will have to gain high confidence in the authenticity of submitted items while learning nothing about them. Proposed inspection systems featuring 'information barriers', designed to hide measurements stored in electronic systems, are at risk of tampering and snooping. Here we show the viability of a fundamentally new approach to nuclear warhead verification that incorporates a zero-knowledge protocol, which is designed in such a way that sensitive information is never measured and so does not need to be hidden. We interrogate submitted items with energetic neutrons, making, in effect, differential measurements of both neutron transmission and emission. Calculations for scenarios in which material is diverted from a test object show that a high degree of discrimination can be achieved while revealing zero information. Our ideas for a physical zero-knowledge system could have applications beyond the context of nuclear disarmament. The proposed technique suggests a way to perform comparisons or computations on personal or confidential data without measuring the data in the first place.

Existing nuclear arms-control agreements between the United States and Russia place limits on the number of deployed strategic nuclear weapons. Verification of these agreements takes advantage of the fact that deployed weapons are associated with unique and easily accountable delivery platforms, that is, missile silos, submarines and strategic bombers, to which agreed numbers of warheads are attributed. The next round of nuclear arms-control agreements, however, may place limits on the total number of nuclear weapons and warheads in the arsenals. This would include tactical weapons as well as deployed and non-deployed weapons. Such agreements would require new verification approaches, including inspections of individual nuclear warheads in storage and warheads entering the dismantlement queue. This is a qualitatively new challenge because the design of nuclear weapons is highly classified information that cannot be exposed to international inspectors. A viable verification approach therefore has to resolve the tension between reliably verifying that the inspected warhead is authentic while avoiding disclosure of information about its design[1–4].

Practitioners and policy makers have been well aware of this conundrum, and prior work by national laboratories in the United States, Russia and the United Kingdom addressed it by using 'information barriers'[2,4]. These barriers consist of sophisticated automated systems that process highly classified information measured during an inspection, but only display results in a yes/no manner. Such systems are inherently complex, and require both parties to trust that they have no 'trapdoors' hidden from the inspector, which could be used to cause a system to declare invalid objects as authentic, nor side channels unknown to the host, which could leak classified information to the inspector or others. These concerns are serious obstacles to adopting such systems.

In this work we consider a fundamentally different approach to this problem. Rather than trying to acquire and analyse classified data behind an engineered information barrier, we use the cryptographic notion of zero-knowledge proofs to ensure that sensitive data are never measured in the first place.

Zero-knowledge proofs (with marbles)

These proofs, invented in the 1980s by Goldwasser, Micali and Rackoff[5], have become an important tool of modern cryptography. They achieve the paradoxical goal of allowing one to prove that a statement is true without revealing why it is true. Such proofs are extremely useful for many digital applications, including privacy-preserving data mining, electronic voting and online auctions[6]. To achieve zero knowledge, Goldwasser *et al.* extended the traditional notion of a proof from a static text to a protocol, which involves randomization and interaction between the prover and verifier. At the end of the protocol, the verifier has a high degree of confidence that the statement is correct, while the prover is guaranteed that the verifier did not learn anything about the data underlying the truth of the statement. For our application, the host submitting warheads for inspection takes the role of the prover and the inspector the role of the verifier.

Whereas classical zero-knowledge proofs are digital protocols, proving statements about mathematical objects, we illustrate the concept using a physical zero-knowledge protocol that is closely related to our proposed verification approach (Fig. 1):

Alice (the host) has two small cups both containing X marbles, where X is some number between 1 and 100. She wants to prove to Bob (the inspector) that both cups contain the same number of marbles, without revealing to him what this number X is. To do so, Alice prepares two buckets, which she claims each contain $(100 - X)$ marbles. Bob now randomly chooses into which bucket which cup is poured. Once this is done, Bob verifies that both buckets contain 100 marbles.

The protocol reveals no information on X because, regardless of the value of X, Bob always sees 100 marbles in both buckets. However, if the cups did not have the same number of marbles, then no matter how Alice prepares the buckets, with a probability of 50% after the pouring at least one of the buckets will not contain 100 marbles. If Alice and Bob repeat this game, say, five times, then if Alice consistently cheats she will be caught with probability $(1 - 2^{-5}) > 95\%$.

[1]Department of Mechanical and Aerospace Engineering, Princeton University, E-Quad, Olden Street, Princeton, New Jersey 08544, USA. [2]Microsoft Research, 1 Memorial Drive, Cambridge, Massachusetts 02142, USA. [3]Princeton Plasma Physics Laboratory, Princeton University, PO Box 451, MS 41, Princeton, New Jersey 08540, USA.

©2014 Macmillan Publishers Limited. All rights reserved

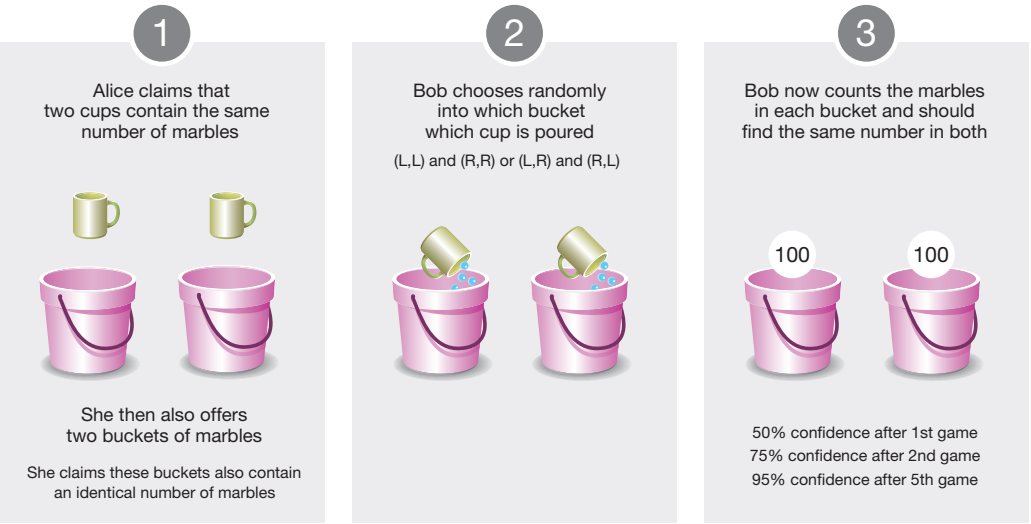


Figure 1 | A zero-knowledge protocol to prove that two cups contain the same number of marbles. L,L indicates left cup into left bucket, and so on. The confidence level increases with the number of games played. See main text for details.

From marbles to neutrons

The relevance of the above protocol to our setting is that we want to show that two or more putative warheads have identical neutron transmission and emission counts under irradiation by high-energy neutrons. We follow the template approach for warhead verification[2], in which a radiation measurement generates a complex and unique fingerprint of an inspected item. This fingerprint is then compared against the fingerprint of one or more templates to confirm that all items are materially identical. Template selection is a critical and challenging step. In practice, templates could be directly selected from deployed weapons so that the inspecting party has high confidence in their authenticity. Strong chain-of-custody measures would have to be in place to assure the inspectors that the templates have not been swapped out between visits. In the case of weapon systems that are not currently deployed, and in other cases where a trusted reference item may not be available, differential measurements could still be valuable. In these cases, measurements of a large number of warheads in a batch could confirm that they are all materially identical. Combined with some other supporting evidence (for example, records confirming the 'pathway' or 'provenance' of at least some of these items), this could provide confidence in the authenticity of all items in the inspected batch[7].

We compare the submitted items by recording the transmission pattern of 14-MeV neutrons, as well as recording the intensity of neutrons emitted at large angles from the items. Active interrogation of nuclear warheads or warhead components with high-energy neutrons and other types of radiation has been successfully demonstrated, but does require detailed safety analyses. Note that, even when exposed to strong neutron sources, the induced fission events in a nuclear warhead produce less than a milliwatt of heat, which is far less than the heat already generated by α-decay and spontaneous fission in items containing, for example, kilogram quantities of plutonium.

Neutron radiographic images of warheads contain highly classified information, but in our case they are actually never measured. Rather, by analogy to the marbles example, they are recorded using detectors that are preloaded with the negative of the radiograph. Preloaded values are not revealed to the inspector. As in the marbles example, after the measurement and if the host is telling the truth, the inspector always sees the same number of counts in every detector. Furthermore, as in the marbles example, preloads supplied with the submitted items are shuffled at random, so if the items actually differ, then no matter how the preloads are chosen, with significant probability the image will not be uniform, and a mismatch will be present on both items.

Unlike the marbles example, neutron measurements are inherently statistically noisy. To avoid conveying information through the noise distribution we use preloaded values that are noisy as well. In particular, since the signal added during interrogation will have a Poisson distribution, we also use a Poisson distribution for the preloads. Using the fact that the sum of two Poisson distributions is also Poisson in character, our protocol achieves the following: the neutron count obtained by any measurement on the template or on any valid submitted item is distributed according to the Poisson distribution with mean equal to a previously agreed-on value, N_{max}, and a standard deviation equal to $(N_{max})^{0.5}$.

Since N_{max} is known in advance by both sides, neither the measurement nor its noise reveals any new information. N_{max} for transmission could reasonably correspond to the maximum number of counts that is expected in the absence of a test item. If a submitted item varies from the true warhead (or the submitted preloads are not identical) an image may be seen that could contain sensitive information. This will be an additional strong incentive for the host not to cheat.

For simplicity of operation, we envision that the host places the detectors for each measurement in a removable board that forms part of the measurement system. Crucially, the inspector chooses which board to use with which test item. As in the marbles example, this means that if the host uses non-identical boards to try to mask invalid items, then with 50% probability the invalidity will be made more evident by the measurement with the mismatched boards. Since we expect that this 'game' will be repeated many times, even a risk-tolerant host would not accept the resulting low chance of success. We note that testing multiple warheads in parallel is an attractive option, because it makes the probability of detecting the use of non-identical preloads significantly higher.

Once the measurements have been completed and the detectors read out, the inspecting party can verify the functionality of the detectors by exposing them to additional neutrons. This is an important advantage of the proposed method. For inspection systems proposed so far, specialists from the inspecting country would not be allowed to examine any equipment once it has seen classified information.

Although we examine here neutron measurements using preloaded non-electronic detectors, there may be other non-electronic zero-knowledge protocols for warhead verification that can avoid the use of engineered information barriers. Indeed, such systems could be complementary to the neutron measurements discussed here.

Monte Carlo analysis

We now show how our approach can be implemented in practice, and that small differences between two objects can be reliably detected. We have analysed the approach with a series of simulations using the general Monte Carlo N-Particle (MCNP) transport code[8]. Construction of a physical experimental set-up is under way.

©2014 Macmillan Publishers Limited. All rights reserved

We propose to use 14-MeV neutrons from a deuterium-tritium neutron generator[9] to interrogate test items, allowing detailed transmission profile measurements and also measurements of neutron intensities at large angles due to elastic and inelastic scattering, fission and $(n,2n)$ reactions. The neutrons from the generator are collimated by 60 cm of polyethylene and illuminate the inspected item (Fig. 2). An array of neutron detectors placed at a distance of 50 cm behind the centre of the item provides the transmission measurements. Additional detectors (not shown) can be positioned with additional shielding at large angles to the beam, that is, in the shadow of the collimator, to measure neutrons emitted from the test item.

Test item

The test item used for this analysis is the unclassified 'British Test Object' (BTO), which consists of concentric rings of polystyrene, tungsten (two rings with a combined mass of 7.74 kg), aluminium, graphite, and steel. The BTO has an outer diameter of 18.9 cm and a height of 5 cm. This test object does not contain special or other nuclear materials, but is used to develop and calibrate imaging systems for diagnostic analysis of nuclear weapons[10]. $(n,2n)$ reactions in the tungsten used in this test object provide a reasonable approximation of induced fission events expected for a nuclear warhead or warhead component. Neutron multiplication in a real item would increase net neutron production rate by some finite amount, but the effect is extremely small for transmission measurements. Furthermore, the energy threshold (~10 MeV) used for the transmission detectors renders them insensitive to fission and $(n,2n)$ neutrons. The BTO is placed in a container in order to avoid revealing to the inspector the appearance or orientation of the inspected item inside the container.

Detector array

To assess the viability of our proposed protocol, we work with a board holding a hexagonal array of 367 detectors consisting of 21 rows of 17 or 18 detectors within an area of about 42 cm × 42 cm. The assumed area of each detector (pixel) is 2 cm^2. By rotating the BTO, the board can image it in any orientation.

In the analysis below, detectors are assumed to be sensitive to neutron energies >10 MeV. Neutrons scattered from the walls of the room are not included in the calculations, but preliminary studies indicate that room return has at most a small impact when using 10-MeV-threshold detectors, particularly if the room is specially prepared for the inspection—for example, with borated polyethylene in front of borated concrete walls.

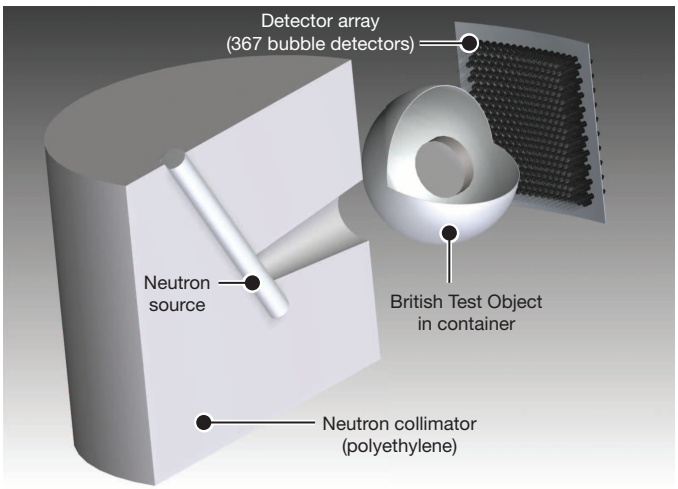


Figure 2 | Experimental set-up with neutron source, neutron collimator, British Test Object in container, and detector array. Large-angle detectors are not shown. See main text for details.

Figure 3 illustrates typical results from a template and a valid item in two different orientations; for reference purposes, the respective neutron radiographs of the test items are shown, but these data are never measured in the inspection, since only preloaded detectors are used. As expected, in the case of inspecting a valid item, detector counts are distributed consistent with a Poisson distribution with mean and variance N_{max}. In the following, only the more challenging side-view orientation of the BTO is used for analyses of four representative diversion scenarios.

Diversion scenarios

To examine diversion scenarios, in which material is removed or replaced, we need to define a decision rule to distinguish passed from failed tests. For our present purposes, we use a very simple rule looking for statistical outliers on predefined groups of pixels. If we denote individual detector counts by the numbers $X_1,\ldots, X_n$, then we can define new numbers $Y_1,\ldots,Y_k$, where every Y_j is the sum of a small number of the X_is, divided by the expected standard deviation of Y for a match case (that is, inspected item identical to template). We define the test to be positive (that is, diversion detected) if there is at least one j with $|Y_j| > T$, where T is a threshold chosen such that in the match case for every j the probability that $|Y_j| > T$ is at most p_{fp}/k where p_{fp} is our allowed false positive rate. Concretely, in our setting, we examine $k = 295$ non-disjoint seven-pixel windows defined by a central detector and its six nearest neighbours. In this case, to achieve a false positive rate $p_{fp} \leq 0.05$, the threshold can be computed numerically to be $T = 3.76$ standard deviations.

Sensitivity of the measurements to diversion scenarios increases with N_{max} and the associated improvements of counting statistics. We therefore examine in the following a series of different diversion scenarios and a range of values for N_{max} to determine system requirements (Table 1). In the full-removal scenario, both tungsten rings are removed from the BTO, which is easily detected even for very low detector counts. Similarly, if lead is used to substitute both tungsten rings, the diversion is clearly distinguishable even by simple visual inspection of the detector board (Fig. 4, top). Our proposed statistical test identifies the diversion in the full-substitution scenario with a probability of true positives, p_{tp}, of > 0.99, even for N_{max} as low as 1,000 detector counts.

The local-removal and local-substitution scenarios are more challenging. In these cases, a 36° sector of the outer tungsten ring is removed or replaced, which corresponds to a diversion of 543 g of tungsten contained in the BTO. To achieve a detection probability of 95%, an N_{max} of 5,000 is required in the case of the localized tungsten removal. When lead is used to substitute for tungsten in the 36° sector, N_{max} increases to 32,000 for the same detection probability.

Note that in these studies no use has been made of the emission detectors at large angles. The more realistic case of substitution of ^{238}U for ^{235}U in a nuclear weapon component results in a reduction by a factor of about two in the induced fission rate due to 14-MeV neutrons. Substitution of reactor-grade for weapon-grade plutonium has a small effect on the directly induced fission rate, but a large effect on the spontaneous fission rate, which could be detected passively by operating the side detectors in the absence of the neutron source. Thus the calculations presented here are conservative.

We note that 5% of the items will be flagged as invalid by our proposed test procedure due to the set 5% false positive rate, even in the case where all items are valid. Retesting flagged items will rapidly determine their validity. If a detection probability for invalid items of 95% is deemed too low, either routine retesting or a greater N_{max} can be implemented to increase this value. The optimization of any retesting scenario, and study of a wider range of host strategies for cheating, as well as inspector strategies for analysing signal patterns to find such cheating, will be the subject of future research.

Preloadable non-electronic detectors

Perhaps the most critical aspect of a viable implementation of the proposed verification approach is the choice of the detector technology. The

©2014 Macmillan Publishers Limited. All rights reserved

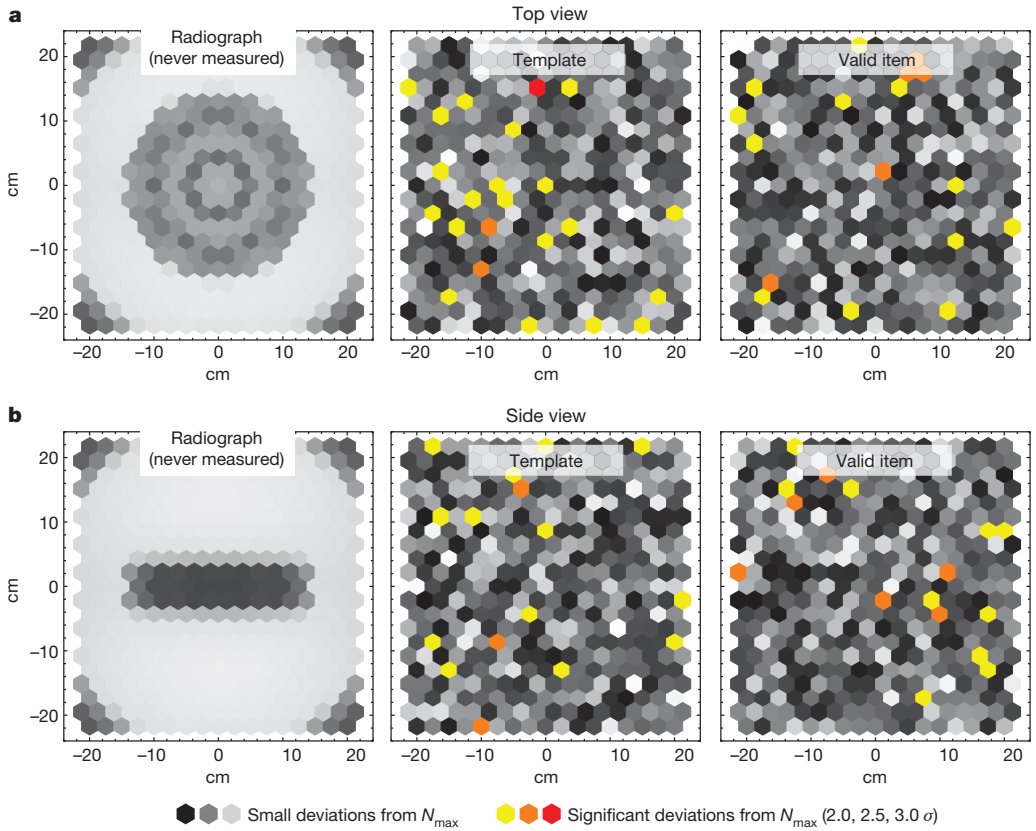


Figure 3 | Results of MCNP5 simulations for interrogations of the British Test Object in two different orientations. Inspection of valid items reveals no information about them. **a**, View from the top; **b**, view from the side. Leftmost panels, radiographs of the items are never measured; they correspond to measurements without preloading the detectors. Other panels visualize total detector counts, including the host's preloads, after the measurements on the template (middle panels) and on an identical (valid) item (rightmost panels), and hence are simply independent Poisson random variables with mean N_{max}. Shades of grey and colours indicate differences from N_{max} scaled to $(N_{max})^{0.5}$. For this calculation, $N_{max} = 5{,}000$.

detectors must have the capability to be preloaded with a desired neutron count before the inspection. At a minimum, this preload has to persist for hours or days and its decay rate, if present, be well characterized. Preloaded counts must be indistinguishable from counts accumulated during irradiation of the test items. Detectors should be energy selective so that the effect of low energy neutrons returning from room walls can be minimized. They should be insensitive to γ-rays, have high efficiency, and permit total counts in the range discussed above. Finally, relying on a non-electronic detection mechanism is highly advantageous given that complex electronic components and circuits are potentially vulnerable to tampering and snooping. We find that at least two detector technologies can meet these criteria: superheated emulsions ('bubble detectors') and neutron activation analysis detectors.

In superheated emulsions, neutron recoil particles trigger the formation of macroscopically observable bubbles from microscopic droplets that are dispersed in an inert matrix[11]. These detectors can be configured to have essentially any desired energy threshold from 10 keV to 10 MeV. Commercially available, polymer-based bubble detectors are limited to a maximum bubble count of the order of a few hundred bubbles, beyond which camera-based imaging techniques cannot resolve bubbles individually. Superheated drop detectors produced with an aqueous gel can be used up to much higher bubble counts. Either optical tomography or magnetic resonance imaging allow the counting of bubbles hidden in the depth of the fluid[12,13]. If the highest N_{max} is desired, multiple detectors can be exposed in series. By the proper choice of a compliant matrix, detectable ageing (growth) of bubbles can be eliminated. Net detection efficiency of the order of 1% can be easily achieved. The emulsions can be contained in opaque containers so that a preload is not visible to the inspector.

For neutron activation analysis, imaging can be undertaken using, for example, an array of hexagonal prisms made of zirconium[14]. ^{90}Zr has a neutron activation threshold of 12 MeV through an $(n,2n)$ reaction. The resulting ^{89}Zr has a half-life of 3.3 days, which must be taken into account to determine the required level of preloading. Counting the γ-rays from ^{89}Zr decay in high-purity germanium well detectors should give a net detection efficiency of about 0.25%. For 3-cm-long prisms, with a cross-sectional area of 2 cm^2, this would provide an N_{max} in the range of 20,000 within one hour, for a commercially available DT neutron generator producing 3×10^8 neutrons s^{-1} (ref. 15). Indium has an appropriate activation response to fission neutrons, with reduced sensitivity in the range of 14 MeV, for use in the side detectors. It will be important to ensure that unshielded, preloaded activation samples are not in the presence of γ-ray detectors before their final exposure in order to avoid providing the inspector with any information about the preloads.

Detectors can be preloaded with counts with the appropriate statistical properties by exposing them to energetic neutrons for a pre-calculated period of time and/or through a pre-calculated depth of shielding. As discussed above, statistical noise in the measurement will not reveal any information. However any systematic measurement errors must be well understood, such that while one detector may be characterized by a different

Table 1 | Probability of an item being flagged as 'invalid' as a function of N_{max} with 10,000 realizations for each case

Item	N_{max}					
	500	1,000	2,000	5,000	10,000	32,000
Valid item*	≤5%	≤5%	≤5%	≤5%	≤5%	≤5%
Full removal of both tungsten rings	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%	>99.9%
Full substitution of tungsten rings by lead	77.7%	99.5%	>99.9%	>99.9%	>99.9%	>99.9%
Local removal of tungsten	Undetectable	15.7%	41.7%	94.6%	>99.9%	>99.9%
Local substitution of tungsten by lead	Undetectable	Undetectable	6.0%	11.7%	30.2%	95.5%

* By design.

©2014 Macmillan Publishers Limited. All rights reserved

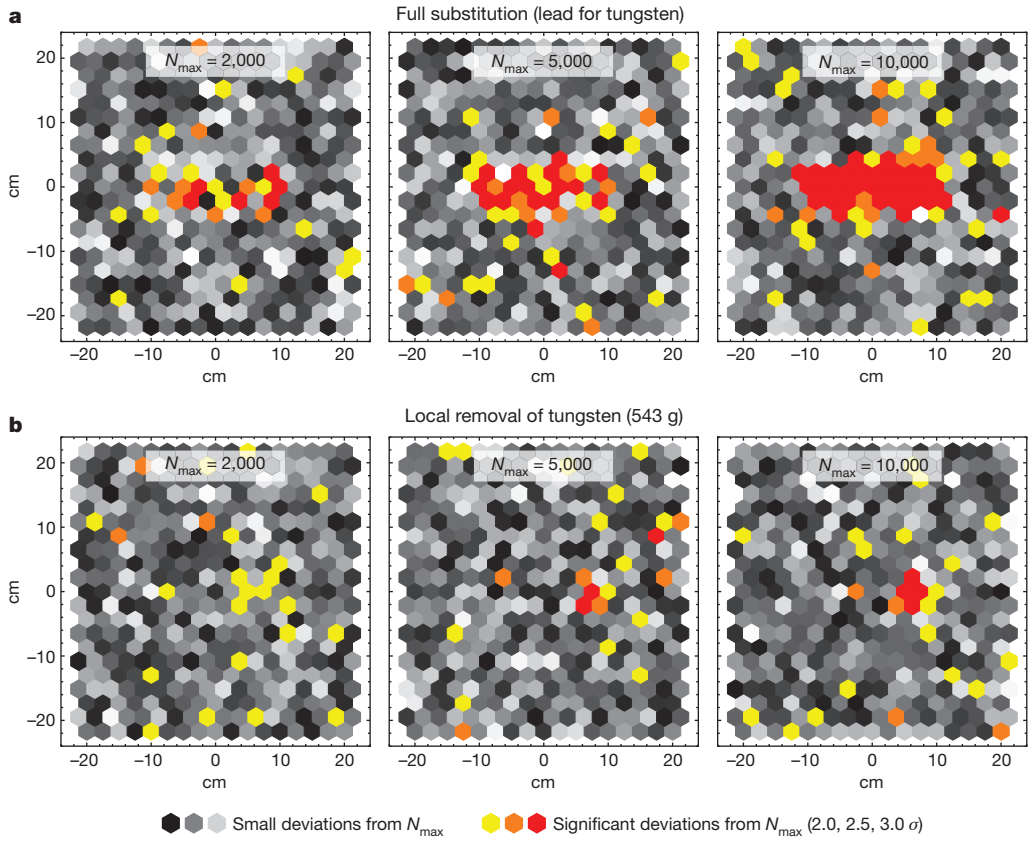


Figure 4 | Results of MCNP5 simulations for two notional diversion scenarios. **a**, Full substitution: tungsten rings in the British Test Object (BTO) are replaced by lead rings. **b**, Local removal: a 36° sector of the outer tungsten ring (mass 543 g) is removed. Inspection of invalid items results in anomalies of the detector counts that become more pronounced with increasing N_{max}. Shades of grey and colours indicate differences from the selected values of N_{max} = 2,000, 5,000 and 10,000 (leftmost, middle and rightmost panels respectively) scaled to $(N_{max})^{0.5}$.

efficiency from another, which can be calibrated out, this efficiency must not vary significantly between the preload and the measurement processes. For example, it is important to maintain control over the temperature of bubble detectors during irradiation. The DT neutron generator must also be well controlled and measurable, so that there is no significant variation in the shape of the neutron field produced nor in the total number of neutrons emitted when irradiating items. An accurate neutron flux monitor can be used to set the irradiation time, so perfect reproducibility is not required in the rate of neutron production. We anticipate that these requirements can be met, but the techniques to achieve the necessary degree of control need to be demonstrated and validated.

The steps following a measurement should be relatively straightforward. Since the information contained in the detectors is in principle unclassified, protocols can be devised that permit using both host-provided and inspector-provided measurement tools. We anticipate that, depending on the strength of the neutron source, the measurements themselves could be completed within hours. Readout would be very quick in the case of bubble detectors, but could take days in the case of some activation detectors. This is not a major constraint, since readout could take place in parallel with other measurements and activities at the site. The authentication process could be accelerated dramatically if N warheads are processed simultaneously (including, for example, $M < N$ reference items). Typically, authentication of one warhead per day can be considered a reasonable target, especially since dismantlement of the warhead itself (including recovery of fissile material and removal of classified features) would take much longer. Authentication is therefore not a significant bottleneck in the process.

Developing a practical inspection system for nuclear warhead verification will be a major undertaking. Ideally, such a system should be jointly developed by partners from weapon and perhaps also non-weapon states. The successful UK–Norway Initiative has shown that such collaborations are possible[16]. Similar efforts could be undertaken for the system proposed here. They would help refine and demonstrate to the satisfaction of all parties the robustness of the method in practical situations, where systematic errors in measurements, small misalignments, or variations in environmental conditions may pose additional challenges that are difficult to anticipate with computer simulations.

More generally, we believe that our approach and techniques could have other applications beyond the area of nuclear disarmament. Once data are measured and converted to digital form, secure comparisons and computations can be performed using many cryptographic tools[17,18]. However, if the measurement device itself cannot be trusted, it is best to ensure the data are never measured in the first place. This could be the case not just for state secrets, but also for personal data, such as biometric data or results of medical tests. For example, following earlier versions of the present work, it was proposed[19] that similar ideas could be used to compare DNA found in a crime scene with a suspect's DNA without actually measuring the latter and creating a DNA profile. Exploring other such applications is an exciting future direction of research on zero-knowledge proofs.

Conclusion

Authenticating nuclear warheads without revealing classified information represents a qualitatively new challenge for international arms-control inspection. Here we have shown an example of a zero-knowledge protocol based on non-electronic differential measurements of transmitted and emitted neutrons that can detect small diversions of heavy metal from a representative test object. This technique will reveal no information about the composition or design of nuclear weapons when only true warheads are submitted for authentication, and so does not require an engineered information barrier. The zero-knowledge approach has the potential to remove a major technical obstacle for verifying deep cuts in the nuclear arsenals, which will probably require verification of individual warheads, rather than whole delivery systems. Timely demonstration of the viability of such an approach could be critical for future rounds of arms-control negotiations.

©2014 Macmillan Publishers Limited. All rights reserved

Received 30 January; accepted 28 April 2014.

1. Comley, C, *et al. Confidence, Security & Verification: The Challenge of Global Nuclear Weapons Arms Control* http://www.fissilematerials.org/library/awe00.pdf (Atomic Weapons Establishment, Aldermaston, UK, 2000).
2. Spears, D. (ed.) *Technology R&D for Arms Control* http://www.fissilematerials.org/library/doe01b.pdf (US Department of Energy, Office of Nonproliferation Research and Engineering, Washington DC, 2001).
3. Fuller, J. in *Cultivating Confidence: Verification, Monitoring, and Enforcement for a World Free of Nuclear Weapons* (ed. Hinderstein, C.) Ch. 4 (Hoover Institution Press, 2010).
4. Anderson, B. *et al. Verification of Nuclear Weapon Dismantlement: Peer Review of the UK MoD Programme* (British Pugwash Group, London, 2012).
5. Goldwasser, S., Micali, S. & Rackoff, C. The knowledge complexity of interactive proof-systems. *SIAM J. Comput.* **18,** 186–208 (1989).
6. Chazelle, B. The security of knowing nothing. *Nature* **446,** 992–993 (2007).
7. Fuller, J. Verification on the road to zero: issues for nuclear warhead dismantlement. *Arms Control Today* **40** (10), 19–27 (December 2010).
8. A general Monte Carlo N-particle (MCNP) transport code. Release MCNP5-1.40 http://mcnp.lanl.gov (Los Alamos National Laboratory, 2005).
9. Roquemore, A. L., Jassby, D. L., Johnson, L. C., Strachan, J. D. & Barnes, C. W. Performance of a 14-MeV neutron generator as an in situ calibration source for TFTR. In *15th IEEE/NPSS Symposium on Fusion Engineering* 114–118 (IEEE, 1993).
10. Hall, J. Uncovering hidden defects with neutrons. *Sci. Technol. Rev.* www.llnl.gov/str/May01/Hall.html, 4–11 (May 2001).
11. d'Errico, F. Radiation dosimetry and spectrometry with superheated emulsions. *Nucl. Instrum. Methods Phys. Res. B* **184,** 229–254 (2001).
12. d'Errico, F., Nath, R., Lamba, M. & Holland, S. K. A position sensitive superheated emulsion chamber for three-dimensional photon dosimetry. *Phys. Med. Biol.* **43,** 1147–1158 (1998).
13. d'Errico, F., Di Fulvio, A., Maryañski, M., Selici, S. & Torrigiani, M. Optical readout of superheated emulsions. *Radiat. Meas.* **43,** 432–436 (2008).
14. Bleuel, D. L. *et al.* Neutron activation analysis at the National Ignition Facility. *Rev. Sci. Instrum.* **83,** 10D313 (2012).
15. Thermo Scientific P 385 Neutron Generator. www.thermoscientific.com/en/product/p-385-neutron-generator.html.
16. Allen, K. *et al.* UK-Norway Initiative (UKNI) approach for the development of a gamma ray attribute measurement system with an integrated information barrier. In *Proc. ESARDA 35th Annual Meeting (Ispra, Italy, June 2013)* (ed. Sevini, F.) (European Commission, 2013).
17. Goldreich, O., Micali, S. & Wigderson, A. in *Proc. 19th Annual ACM Conference on Theory of Computing* (ed. Aho, A. V.) 218–229 (ACM Press, 1987).
18. Lindell, Y. & Pinkas, B. Secure multiparty computation for privacy-preserving data mining. *J. Privacy Confident.* **1,** 59–98 (2009).
19. Fisch, B., Freund, D. & Naor, M. Physical zero-knowledge proofs of physical properties. In *Proceedings of CRYPTO 2014* (eds Garay, J. & Gennaro, R.) (in the press); http://www.iacr.org/conferences/crypto2014/acceptedpapers.html.

Acknowledgements This project was supported by Global Zero and the US Department of State, the Princeton Plasma Physics Laboratory (DOE contract DE-AC02-09CH11466) and in-kind contributions from Microsoft Research New England. We thank D. Dobkin, F. d'Errico, J. Fuller, D. MacArthur, J. Mihalczo, S. Philippe and M. Walker for discussions and feedback. We thank S. Philippe (Princeton University) for graphics in Fig. 2. All simulations were run on Princeton University's High Performance Cluster.

Author Contributions A.G. was project leader and performed all of the MCNP5 calculations. B.B. and R.J.G. contributed, along with A.G., to developing the overall approach and the application of zero-knowledge proofs to warhead verification.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.G. (aglaser@princeton.edu).

©2014 **Macmillan Publishers Limited. All rights reserved**

ARTICLE

doi:10.1038/nature13445

Aspergillomarasmine A overcomes metallo-β-lactamase antibiotic resistance

Andrew M. King[1,2], Sarah A. Reid-Yu[3], Wenliang Wang[1], Dustin T. King[4], Gianfranco De Pascale[1], Natalie C. Strynadka[4], Timothy R. Walsh[5], Brian K. Coombes[3] & Gerard D. Wright[1,2,3]

The emergence and spread of carbapenem-resistant Gram-negative pathogens is a global public health problem. The acquisition of metallo-β-lactamases (MBLs) such as NDM-1 is a principle contributor to the emergence of carbapenem-resistant Gram-negative pathogens that threatens the use of penicillin, cephalosporin and carbapenem antibiotics to treat infections. To date, a clinical inhibitor of MBLs that could reverse resistance and re-sensitize resistant Gram-negative pathogens to carbapenems has not been found. Here we have identified a fungal natural product, aspergillomarasmine A (AMA), that is a rapid and potent inhibitor of the NDM-1 enzyme and another clinically relevant MBL, VIM-2. AMA also fully restored the activity of meropenem against Enterobacteriaceae, *Acinetobacter* spp. and *Pseudomonas* spp. possessing either VIM or NDM-type alleles. In mice infected with NDM-1-expressing *Klebsiella pneumoniae*, AMA efficiently restored meropenem activity, demonstrating that a combination of AMA and a carbapenem antibiotic has therapeutic potential to address the clinical challenge of MBL-positive carbapenem-resistant Gram-negative pathogens.

The β-lactams (penicillins, cephalosporins, carbapenems and monobactams) are one of the most important and frequently used classes of antibiotics in medicine and are essential in the treatment of serious Gram-negative infections. Since the clinical introduction of penicillins and cephalosporins over 60 years ago, the emergence of β-lactamases, enzymes that hydrolyse the β-lactam ring that is essential for the cell-killing activity of the antibiotics, has been an ongoing clinical problem[1]. Antibiotic resistance has intensified medicinal chemistry efforts to broaden antibacterial spectrum while shielding the core β-lactam scaffold from β-lactamase-catalysed hydrolysis. The result has been multiple generations of β-lactams with improved efficacy and tolerance to existing β-lactamases. However, pathogenic bacteria have in turn evolved further resistance mechanisms primarily by acquiring new or modified β-lactamases. This is typified by the emergence of extended spectrum β-lactamases that inactivate many of the latest generation cephalosporins and penicillins (but not carbapenems)[2]. Consequently, the past two decades have seen substantial increases in the utilization of carbapenems such as imipenem and meropenem. Predictably, this increase in carbapenem consumption has been accompanied by the emergence of carbapenem-resistant Gram-negative pathogens (CRGNP)[3,4]. In particular, carbapenem-resistant Enterobacteriaceae (CRE) is a growing crisis across the globe[5] as witnessed by recent outbreaks in Chicago[6] and British Columbia[7].

MBLs are an unmet clinical challenge

Carbapenemases, β-lactamases that inactivate carbapenems, can be divided into two categories based on their mechanism of β-lactam ring hydrolysis. The first use an active site serine residue that covalently attacks the β-lactam ring for example, KPC and OXA-48 types[8]. The second are MBLs that use Zn^{2+} atoms to activate a nucleophilic water molecule that opens the ring, for example, VIM and NDM types[9]. Several inhibitors of serine β-lactamases are clinically available as co-drugs in which the inhibitor is formulated with a β-lactam antibiotic in order to overcome resistance (for example, clavulanic acid with amoxicillin, tazobactam with piperacillin, sulbactam with ampicillin and the more recent serine β-lactamase inhibitor avibactam, which is in phase III clinical trials paired with various cephalosporins)[10]. Despite ongoing efforts[11,12], there are no equivalent inhibitors for MBLs in the clinic for practical and technical reasons. First, until recently, MBL-derived CRE was not thought to be a major clinical problem and its rapid increase has outpaced MBL-inhibitor development. Second, the development of a single inhibitor to neutralise key clinically important MBLs (VIM and NDM) has been deemed too technically challenging, and overcoming *in vivo* toxicity associated with cross reactivity with human metalloenzymes has been a concern. With the recent emergence of MBLs as a major clinical threat, a potent and safe inhibitor of MBLs particularly against VIM and NDM would greatly benefit infectious disease management.

Cell-based screen for inhibitors of NDM-1

We initiated a cell-based screen for inhibitors of the NDM-1 MBL using our in-house collection of dimethylsulphoxide (DMSO)-dissolved natural product extracts derived from environmental microorganisms. To increase the sensitivity of the screen to MBL inhibitor discovery we generated a test strain of *Escherichia coli* BW25113 in which the *bamB* and *tolC* genes were independently deleted (*E. coli* BW25113 *ΔbamBΔtolC*). BamB is essential to outer membrane porin assembly and disruption results in increased permeability to small molecules[13] and TolC is a key component of tripartite small molecule efflux systems such as AcrA-AcrB-TolC that serve to actively eliminate small molecules from the cell[14]. Therefore, *E. coli* BW25113*ΔbamBΔtolC* increases the sensitivity of the screen to discovery of hits. This strain was further modified by integrating the bla_{NDM-1} β-lactamase gene under control of the *pLac* promoter into the chromosome (*E. coli* BW25113 *ΔbamBΔtolCΔaraDAB::pLac*(bla_{NDM-1})). We screened this strain in the presence of a sub-lethal concentration of meropenem (one quarter of the minimum inhibitory concentration 0.125 μg ml^{-1}) in combination with ~500 natural product extracts.

The screen generated one reproducible hit from an extract of a strain of *Aspergillus versicolor* (as identified by 18S rRNA gene sequence) with excellent ability to restore meropenem antibiotic activity against the

[1]M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [2]Department of Chemistry and Chemical Biology, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [3]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L8S 4K1, Canada. [4]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada. [5]Cardiff Institute of Infection & Immunity, Cardiff University, Cardiff CF14 4XN, UK.

©2014 Macmillan Publishers Limited. All rights reserved

E. coli screening strain. The selectivity of the extract to neutralise NDM-1 activity *in vitro* was confirmed using purified enzyme and the colorimetric β-lactamase substrate nitrocefin and with the carbapenem drug meropenem. Activity-guided purification of the active compound from the fermentation broth of *A. versicolor* and subsequent detailed chemical characterization identified the MBL inhibitor as aspergillomarasmine A (AMA; Fig. 1a, Extended Data Table 1 and Extended Data Figs 1–5), a fungus-derived molecule that was discovered and reported in the early 1960s for its wilting and necrotic activity on plant leaves[15]. This molecule was re-evaluated in the 1980s as an inhibitor of angiotensin-converting enzyme (ACE)[16] and in the early 1990s as a preclinical candidate for the inhibition of activation of human endothelin[17], a peptide that modulates blood vessel muscle contraction. The activation of this vasoconstricting peptide requires endothelin-converting enzyme, which, like angiotensin-converting enzyme, is a metalloproteinase that shares some mechanistic similarities with MBLs. This previous work demonstrated that AMA was well-tolerated and had low toxicity in mice (dose lethal to 50% of animals tested (LD_{50}) was 159.8 mg per kg, i.v. compared to EDTA at 28.5 mg per kg) and had no effect on mean atrial blood pressure[18].

AMA is a potent inactivator of MBLs

AMA showed potent *in vitro* dose-dependent inhibition of NDM-1 and the related MBL VIM-2 (Fig. 1b), with weaker activity against the IMP-7 MBL. AMA had no effect on the serine β-lactamases TEM-1 and CTX-M-15 as well as the serine carbapenemases KPC-2 and OXA-48 (Fig. 1b and Extended Data Fig. 6b). The fact that AMA was identified as an inhibitor of unrelated mammalian metalloproteinases in the past suggested that the compound interacts with the MBL metal centres. Inhibition of NDM-1 was shown to be irreversible after removal of AMA by gel filtration (Fig. 1c), but enzymatic activity could be restored by addition of excess $ZnSO_4$, consistent with a metal depletion mechanism (Fig. 1d). The inhibition of mammalian metalloenyzmes could be seen as a potential side effect, however, AMA was only able to reduce the activity of rabbit lung ACE by ~ 35% in concentration–response assays (Extended Data Fig. 6a). Extended incubation of the metalloenzymes NDM-1, VIM-2, IMP-7 and ACE at high concentrations of AMA (0.5 mM) in Zn^{2+}-depleted buffer prepared as previously described[19] led to complete inactivation of NDM-1 and VIM-2, and ~70% and ~50% inhibition of activity in IMP-7 and ACE, respectively, demonstrating selectivity towards NDM and VIM MBLs (Extended Data Table 2). Time-dependent inactivation was shown to be saturable for NDM-1 ($K_i = 11$ nM, $k_{+2} = 0.0062\ s^{-1}$) and VIM-2 ($K_i = 7$ nM, $k_{+2} = 0.0065\ s^{-1}$) (Fig. 1e), consistent with an inactivation mechanism whereby AMA removes Zn^{2+} ions similar to known *in vitro* chelators that interact with subclass B1 MBLs[20]. This mechanism of action was confirmed by inductively coupled mass spectrometry that showed a loss of ~ 1.8 Zn equivalents in NDM-1 inactivated by AMA (Fig. 1f).

Systematic titration of AMA and meropenem concentrations against our engineered *E. coli* and a panel of clinical CRE strains demonstrated that AMA restored meropenem activity consistent with NDM inhibition. Checkerboard MIC studies confirmed the expected synergy between meropenem and AMA only in NDM-1-expressing CRE and not in carbapenem-sensitive strains (Fig. 2a, b). Fractional inhibitory concentration (FIC) index values were determined to be < 0.1 for a panel of 16 clinical CRE isolates tested against meropenem and AMA combinations (FIC values of ≤ 0.5 are defined as synergistic[21]) (Extended Data Table 3). Potentiation of AMA (8 μg ml^{-1}) with meropenem was further investigated using 229 MBL positive (SPM-1, IMP, NDM, AIM and VIM) non-clonal clinical isolates (Enterobacteriaceae, *Acinetobacter* spp. and *Pseudomonas* spp.) (Fig. 2c). A total of 76 isolates were also tested that possessed serine carbapenemases, or MBLs and serine carbapenemases. Strains were amassed over a 10 year period as part of a global MBL collection including isolates from Russia, India, Pakistan, Australia, North Africa and South America. AMA restored meropenem sensitivity (2 μg ml^{-1}) in 88% of NDM positives isolates and 90% of VIM positive isolates. Importantly, AMA was active in *Pseudomonas* spp. (mainly *Pseudomonas aeruginosa*), which are viewed as a highly challenging model for new antibiotics. AMA showed very little potentiation with SPM-1, IMP and AIM but these MBLs are less numerous than the global VIM and NDM MBLs and therefore deemed to be less clinically relevant. The lack of potentiation with IMP-expressing strains correlates well with biochemical data showing less potent inactivation of purified IMP-7 compared to NDM-1 or VIM-2 (Extended Data Table 2).

AMA reverses carbapenem resistance *in vivo*

The resistance profile of NDM-1-positive clinical CRE and the efficacy with which AMA potentiated meropenem activity against NDM-1-positive clinical CRE suggested that AMA would reverse NDM-1-mediated resistance to meropenem *in vivo* and restore clinical efficacy of this antibiotic. To test this, CD1 mice were infected intraperitoneally with a lethal dose of NDM-1-positive *K. pneumoniae* N11-2218

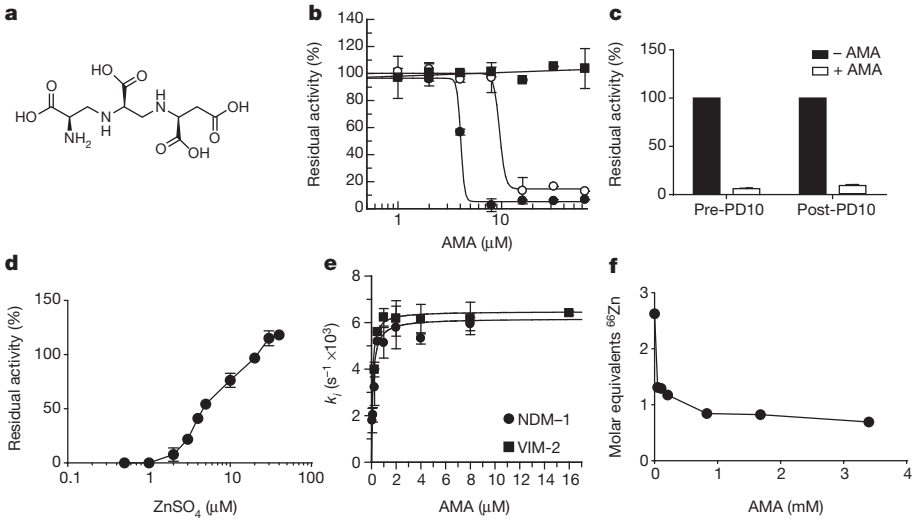


Figure 1 | AMA inactivates MBLs. **a**, Chemical structure of AMA. **b**, AMA inhibits NDM-1 (black circles) (IC_{50} 4.0 ± 1.0 μM) and VIM-2 (white circles) (IC_{50} of 9.6 ± 2.4 μM). Activity of OXA-48 (black squares) was unaffected by AMA. **c**, Removal of AMA via PD10 column does not restore NDM-1 activity, confirming irreversible inactivation. **d**, Addition of excess $ZnSO_4$ restores activity post-inactivation. **e**, The rate of inactivation of NDM-1 and VIM-2 is saturable with AMA concentration. **f**, ICP-MS confirms depletion of Zn from NDM-1. For all experiments error bars denote standard deviation of three technical replicates.

©2014 Macmillan Publishers Limited. All rights reserved

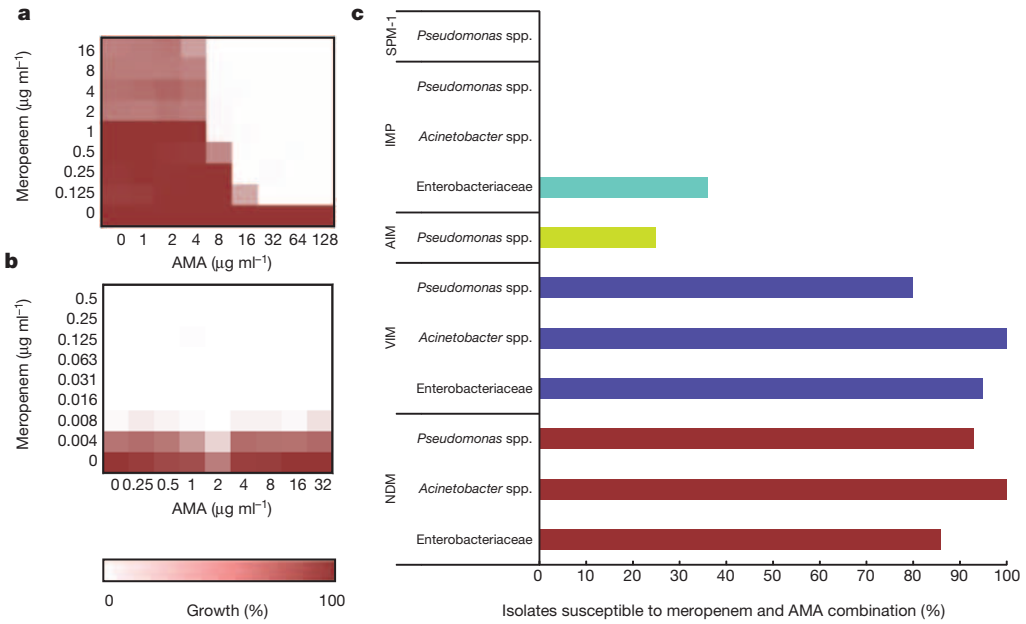


Figure 2 | AMA potentiates the activity of meropenem against carbapenem-resistant Gram-negative pathogens. **a**, **b**, Microdilution checkerboard analysis showing the combined effect of AMA and meropenem selectively against CRE (**a**, *K. pneumoniae* N11-2218 MIC meropenem = 32 μg ml^{-1}) but not a carbapenem-sensitive strain (**b**, *E. coli* BW25113 MIC = 0.008–0.016 μg ml^{-1}). Heat plots shows the average of two technical replicates. **c**, VIM- and NDM-expressing Gram-negative pathogens were highly susceptible to meropenem and AMA combination (2 μg ml^{-1} and 8 μg ml^{-1}, respectively) whereas AIM-, IMP-, and SPM-1-expressing isolates remained resistant.

to initiate a lethal systemic infection and the effects of meropenem or AMA monotherapy or antibiotic-inhibitor combination therapy was evaluated. Preliminary dosing experiments determined empirically that the bacterial load of NDM-1-positive *K. pneumoniae* in tissues was unaffected by treatment with AMA alone, and that this strain was resistant to meropenem monotherapy at doses below 50 mg per kg, leading to lethal infection (Extended Data Fig. 7). However, combination therapy with AMA and meropenem significantly reduced the bacterial load in the spleen (Fig. 3a) and to a lesser extent in the liver (Fig. 3b) after a single intraperitoneal dose. Remarkably, although meropenem or AMA alone were unable to prevent lethal infection by NDM-1-positive *K. pneumoniae*, a single dose of combination therapy led to > 95% survival at 5 days following infection (Fig. 3c).

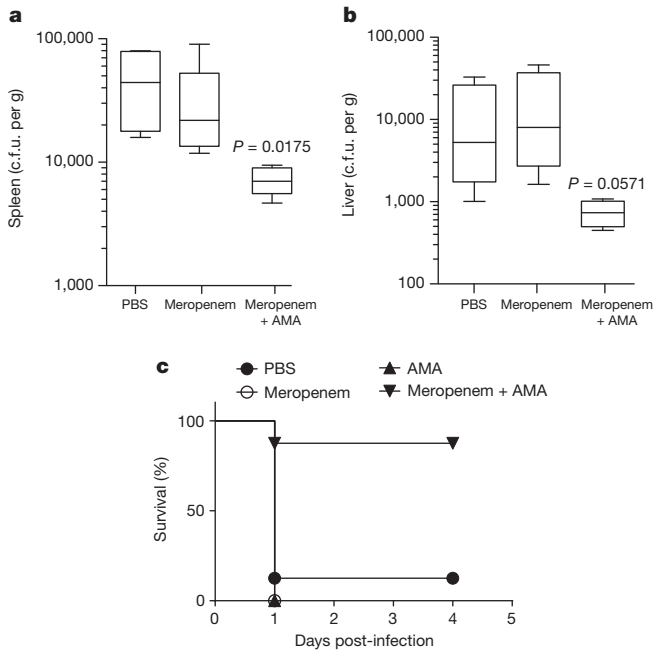


Figure 3 | AMA rescues meropenem activity *in vivo*. CD-1 mice were given a sub-lethal dose of *K. pneumoniae* N11-2218 (meropenem MIC 32 μg ml^{-1}) by i.p. injection. **a**, **b**, Groups of mice were treated with a single dose of meropenem (10 mg per kg), a combination of meropenem (10 mg per kg) plus AMA (10 mg per kg), or PBS by subcutaneous injection. Bacterial load in the spleen (**a**) and liver (**b**) was determined by selective plating. Data are the means with standard error from two separate experiments ($n = 7$ per group). **c**, For survival experiments, CD-1 mice were given a lethal dose of *K. pneumoniae* N11-2218, and treated with a single dose of meropenem (10 mg per kg), a combination of meropenem (10 mg per kg) plus AMA (30 mg per kg), AMA alone (30 mg per kg), or PBS by subcutaneous injection. Data are the means with standard error from four separate experiments ($n = 12$ per group, except the AMA only treatment in which $n = 13$). Groups were analysed using a non-parametric Mann–Whitney *U*-test. *P* values < 0.05 were considered statistically significant.

AMA presents a non-toxic candidate for an antibiotic adjuvant that can overcome resistance mediated by NDM and VIM MBLs and re-sensitize carbapenem-resistant Gram-negative pathogens to carbapenems. Active drug/inhibitor combinations continue to be highly successful in the clinic with inhibitors targeted to serine β-lactamases[22]. AMA presents for the first time, *in vitro* and *in vivo*, complementary activity against key MBLs that have become rapidly global and result in significant human morbidity particularly in developing countries. In combination with a β-lactam antibiotic such as meropenem as we show here, resistance can be overcome and antibiotic activity fully restored. AMA (or semi-synthetic derivatives) is therefore an excellent lead for an antibiotic adjuvant co-therapy to address the recent emergence of MBLs in the clinic.

METHODS SUMMARY

E. coli strain BW25113 was modified by deletion of the *bamB* and *tolC* genes so as to increase permeability to and reduce efflux of small molecules. This strain was then further modified by the single-copy chromosomal insertion of the bla_{NDM-1} gene under control of the *pLac* promoter. Microbial natural product extracts were screened against this strain in the presence of sublethal one-quarter MIC meropenem. AMA was purified from a hit extract and its structure elucidated by NMR, mass spectrometry and polarimetry. Half-maximum inhibitory concentration (IC_{50}) values for AMA were determined using nitrocefin (for MBLs and Ser BLs) and furanacryloyl-L-phenylalanylglycylglycine (for ACE)[23] as reporter substrates against the following purified enzymes: MBLs IMP-7, VIM-2, and NDM-1; SBLs CTX-M-15, KPC-2, OXA-48 and TEM-1; as well as ACE from rabbit lung. IC_{50}, reversibility, Zn^{2+} restoration, and inactivation enzyme assays were performed in 50 mM HEPES pH 7.5 and measured using a SpectraMax reader (Molecular Devices). ICP-MS experiments were conducted using purified NDM-1 at 5 mg ml^{-1} and varying concentrations of AMA with subsequent dilution and transfer by nebulization into a NexION 3000 ICP mass spectrometer (Perkin Elmer). FIC values were determined using standard methods[21]. Various concentrations of AMA were tested in combination with 2 mg l^{-1} meropenem against over 200 MBL-expressing clinical isolates including *Pseudomonas* spp., *Acinetobacter*

©2014 Macmillan Publishers Limited. All rights reserved

spp., and Enterobacteriaceae. A dose of 2×10^6 colony-forming units (c.f.u.) of *Klebsiella pneumoniae* N11-2218 was used for all organ bacterial load experiments and a dose of 5×10^7 c.f.u. for all survival experiments. For all experiments, mice were treated with compound 30 min post-infection.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 February; accepted 7 May 2014.

1. Frère, J. M. *Beta-Lactamases* (Nova Science, 2011).
2. Pitout, J. D. & Laupland, K. B. Extended-spectrum β-lactamase-producing Enterobacteriaceae: an emerging public-health concern. *Lancet Infect. Dis.* **8,** 159–166 (2008).
3. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States* (Atlanta, Georgia, 2013).
4. Edelstein, M. V. *et al.* Spread of extensively resistant VIM-2-positive ST235 *Pseudomonas aeruginosa* in Belarus, Kazakhstan, and Russia: a longitudinal epidemiological and clinical study. *Lancet Infect. Dis.* **13,** 867–876 (2013).
5. Patel, G. & Bonomo, R. A. "Stormy waters ahead": global emergence of carbapenemases. *Front. Microbiol.* **4,** 48 (2013).
6. Frias, M. *et al.* New Delhi metallo-β-lactamase-producing *Escherichia coli* associated with endoscopic retrograde cholangiopancreatography — Illinois, 2013. *MMWR Morb. Mortal. Wkly Rep.* **62,** 1051 (2014).
7. Chang, Y. Laboratory Trends, December 17. 8 (BC Public Health Microbiology & Reference Laboratory, Vancouver, British Columbia, 2013).
8. Yigit, H. *et al.* Novel carbapenem-hydrolyzing beta-lactamase, KPC-1, from a carbapenem-resistant strain of *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* **45,** 1151–1161 (2001).
9. Bush, K. Proliferation and significance of clinically relevant β-lactamases. *Ann. NY Acad. Sci.* **1277,** 84–90 (2013).
10. Drawz, S. M. & Bonomo, R. A. Three decades of β-lactamase inhibitors. *Clin. Microbiol. Rev.* **23,** 160–201 (2010).
11. Fast, W. & Sutton, L. D. Metallo-β-lactamase: inhibitors and reporter substrates. *Biochim. Biophys. Acta* **1834,** 1648–1659 (2013).
12. Buynak, J. D. β-Lactamase inhibitors: a review of the patent literature (2010–2013). *Expert Opin. Ther. Pat.* **23,** 1469–1481 (2013).
13. Ricci, D. P. & Silhavy, T. J. The Bam machine: a molecular cooper. *Biochim. Biophys. Acta* **1818,** 1067–1084 (2012).
14. Blair, J. M. & Piddock, L. J. Structure, function and inhibition of RND efflux pumps in Gram-negative bacteria: an update. *Curr. Opin. Microbiol.* **12,** 512–519 (2009).
15. Haenni, A. L. *et al.* Structure chimique des aspergillomarasmines A et B. *Helv. Chim. Acta* **48,** 729–750 (1965).
16. Mikami, Y. & Suzuki, T. Novel microbial inhibitors of angiotensin-converting enzyme, aspergillomarasmines A and B. *Agric. Biol. Chem.* **47,** 2693–2695 (1983).
17. Arai, K. *et al.* Aspergillomarasmine A and B, potent microbial inhibitors of endothelin-converting enzyme. *Biosci. Biotechnol. Biochem.* **57,** 1944–1945 (1993).
18. Matsuura, A. *et al.* Pharmacological profiles of aspergillomarasmines as endothelin converting enzyme inhibitors. *Jpn. J. Pharmacol.* **63,** 187–193 (1993).
19. Hernandez Valladares, M. *et al.* Zn(II) dependence of the *Aeromonas hydrophila* AE036 metallo-β-lactamase activity and stability. *Biochemistry* **36,** 11534–11541 (1997).
20. Docquier, J. D. *et al.* On functional and structural heterogeneity of VIM-type metallo-beta-lactamases. *J. Antimicrob. Chemother.* **51,** 257–266 (2003).
21. Pillai, S. K., Moellering, R. C. Jr & Eliopoulos, G. M. in *Antibiotics in Laboratory Medicine* (ed. V. Lorian) 365–440 (Williams & Wilkins, 2005).
22. Shlaes, D. M. New β-lactam-β-lactamase inhibitor combinations in clinical development. *Ann. NY Acad. Sci.* **1277,** 105–114 (2013).
23. Holmquist, B., Bunning, P. & Riordan, J. F. A continuous spectrophotometric assay for angiotensin converting enzyme. *Anal. Biochem.* **95,** 540–548 (1979).

Acknowledgements We thank M. Mulvey (Public Health Agency of Canada) and R. Melano (Public Health Ontario) for clinical strains. We thank L. Rossi for her work in constructing the screening strain. AMA inhibition activity on clinical strains by T.R.W. was funded by the UK Medical Research Council (G1100135). This research was funded by a Canadian Institutes of Health Research Grant (MT-13536), Natural Sciences and Engineering Research Council Grant (237480), and by a Canada Research Chair in Infectious Disease Pathogenesis (to B.K.C.) and Antibiotic Biochemistry (to G.D.W.).

Author Contributions A.M.K., G.D.W., S.A.R.-Y., B.K.C., T.R.W. and N.C.S. designed experiments; G.D.P. and A.M.K. designed and engineered the *E. coli* strains and screened extracts; A.M.K. synthesized nitrocefin; A.M.K. cloned constructs; A.M.K. and D.T.K. purified enzymes; A.M.K. performed enzyme kinetics; W.W. and A.M.K. fermented WAC-138 and purified AMA; W.W. elucidated AMA structure; A.M.K. performed FIC experiments; D.T.K. performed ICP-MS; S.A.R.-Y. and B.K.C. designed the animal studies; S.A.R.-Y. and A.M.K. performed animal experiments; T.R.W. performed the clinical isolate screen; and A.M.K. and G.D.W. principally wrote the manuscript with input from all.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.D.W. (wrightge@mcmaster.ca).

©2014 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature13440

The selective tRNA aminoacylation mechanism based on a single G•U pair

Masahiro Naganuma[1,2,3], Shun-ichi Sekine[1,2,4], Yeeting Esther Chong[5]†, Min Guo[5]†, Xiang-Lei Yang[5], Howard Gamper[6], Ya-Ming Hou[6], Paul Schimmel[5,7] & Shigeyuki Yokoyama[1,2,3]

Ligation of tRNAs with their cognate amino acids, by aminoacyl–tRNA synthetases, establishes the genetic code. Throughout evolution, $tRNA^{Ala}$ selection by alanyl–tRNA synthetase (AlaRS) has depended predominantly on a single wobble base pair in the acceptor stem, G3•U70, mainly on the k_{cat} level. Here we report the crystal structures of an archaeal AlaRS in complex with $tRNA^{Ala}$ with G3•U70 and its A3•U70 variant. AlaRS interacts with both the minor- and the major-groove sides of G3•U70, widening the major groove. The geometry difference between G3•U70 and A3•U70 is transmitted along the acceptor stem to the 3′-CCA region. Thus, the 3′-CCA region of $tRNA^{Ala}$ with G3•U70 is oriented to the reactive route that reaches the active site, whereas that of the A3•U70 variant is folded back into the non-reactive route. This novel mechanism enables the single wobble pair to dominantly determine the specificity of tRNA selection, by an approximate 100-fold difference in k_{cat}.

The genetic code requires extremely high aminoacyl-tRNA synthesis specificity. Each aminoacyl-tRNA synthetase (aaRS) ligates its specific amino acid to the adenosine in the conserved 3′-terminal CCA sequence of tRNA. The identity, or the amino acid-accepting specificity, of each tRNA is determined by a particular set of structural features (identity set)[1,2]. The identity sets of most tRNAs consist of a limited number of nucleotides, such as anticodon nucleotides[3–6].

The discovery that the identity set of alanine tRNA ($tRNA^{Ala}$) from *Escherichia coli* is just the single G•U wobble base pair, in the middle of the amino acid-acceptor stem (G3•U70)[7,8], engendered the concept of the second genetic code[9]. A G3•U70 to Watson–Crick mutation of *E. coli* $tRNA^{Ala}$ and reverse mutations of non-alanine tRNAs drastically decreased and increased, respectively, the aminoacylation by alanyl-tRNA synthetase (AlaRS)[7,8], reportedly on the k_{cat} level[10]. Throughout evolution from bacteria to humans, AlaRS recognizes G3•U70 of $tRNA^{Ala}$ as the dominant identity determinant[11] (Fig. 1a, b).

With a specific dependence on G3•U70, *E. coli* AlaRS alanylates small RNAs recapitulating the $tRNA^{Ala}$ acceptor stem (7–12 base pair stem-and-loop or duplex RNAs)[1,7,12–17]. These model RNAs were used for analyses with nucleotide analogues, such as inosine and deoxyribonucleotides, to elucidate which functional groups in the $tRNA^{Ala}$ acceptor stem are important for alanylation. First, unlike the Watson–Crick pairs, G and U in the G•U pair (Fig. 1a) are shifted towards the minor and major groove sides, respectively, and form two hydrogen bonds without using the 2-amino group of G and the 4-carbonyl group of U[18]. The non-hydrogen-bonded nature of the 2-amino group in the G3•U70 wobble pair is important for aminoacylation[19]. Furthermore, the 2′-hydroxyl groups of U70 and C71 and the base of A73 contribute to alanylation. It is likely, therefore, that a number of nucleotides in the acceptor stem are involved, directly or indirectly, in interactions with AlaRS. However, it is still mysterious how just a single pair at positions 3•70 plays the dominant role among the interacting nucleotides, to determine the strict specificity for $tRNA^{Ala}$ (refs 7, 8, 20).

Here we report two crystal structures of AlaRS from the archaeon *Archaeoglobus fulgidus*, in complexes with $tRNA^{Ala}$ bearing G3•U70 ($tRNA^{Ala}$/GU) and with a variant bearing A3•U70 ($tRNA^{Ala}$/AU), together with an alanyl-AMP analogue. The tRNA recognition domain of AlaRS widens the major groove, and thereby contacts both the minor- and major-grooves of the acceptor stem. Between the AlaRS•$tRNA^{Ala}$/GU and AlaRS•$tRNA^{Ala}$/AU structures, the local geometrical difference between G3•U70 and A3•U70 (Fig. 1a) propagates into the drastically distinct orientation of the CCA region, so that the 3′-end A76 of $tRNA^{Ala}$/AU is kept far from the aminoacylation active site. This unprecedented mechanism explains why just a small difference in the sequence (the single base pair at positions 3 and 70) results in the specific selection of $tRNA^{Ala}$, predominantly on the k_{cat} level.

The AlaRS•$tRNA^{Ala}$ complex structures

The crystal structure of the complex of *A. fulgidus* AlaRS, $tRNA^{Ala}$, and an alanyl-adenylate analogue, 5′-*O*-[*N*-(L-alanyl)sulfamoyl] adenosine (Ala-SA)[21], was solved at 3.3 Å resolution (Fig. 1c–e and Extended Data Table 1). Furthermore, we determined the 3.5 Å-resolution structure of a non-reactive complex of AlaRS with a $tRNA^{Ala}$ variant possessing A3•U70 instead of G3•U70 ($tRNA^{Ala}$/AU). In both structures, the crystallographic asymmetric unit contains one homodimer of *A. fulgidus* AlaRS, which is consistent with the results from gel-filtration chromatography of *A. fulgidus* AlaRS[22] and sedimentation equilibrium analyses of *A. fulgidus* and *E. coli* AlaRSs (ref. 23, Methods).

In both structures, AlaRS is composed of the aminoacylation, tRNA-recognition, editing, and C-terminal domains, among which the latter three consist further of the Mid1/Mid2, β-barrel/editing-core, and helical/globular subdomains, respectively[24–31] (Fig. 1c, d). The C-terminal and editing domains constitute the dimer interface (Fig. 1d, e). Both aminoacylation domains of the dimer bind Ala-SA, whereas only subunit A binds $tRNA^{Ala}$ using its tRNA-recognition and C-terminal domains. Isothermal titration calorimetry revealed that one AlaRS dimer binds

[1]RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [2]Department of Biophysics and Biochemistry and Laboratory of Structural Biology, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [3]RIKEN Structural Biology Laboratory, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [4]Division of Structural and Synthetic Biology, RIKEN Center for Life Science Technologies, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [5]The Skaggs Institute for Chemical Biology and the Department of Cell and Molecular Biology, The Scripps Research Institute, BCC-379, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [6]Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA. [7]The Scripps Florida Research Institute, 130 Scripps Way, 3B3 Jupiter, Florida 33458-5284, USA. †Present addresses: aTyr Pharma, 3545 John Hopkins Court, San Diego, California 92121, USA (Y.E.C.); Department of Cancer Biology, The Scripps Research Institute, 130 Scripps Way, Jupiter, Florida 33458, USA (M.G.).

©2014 Macmillan Publishers Limited. All rights reserved

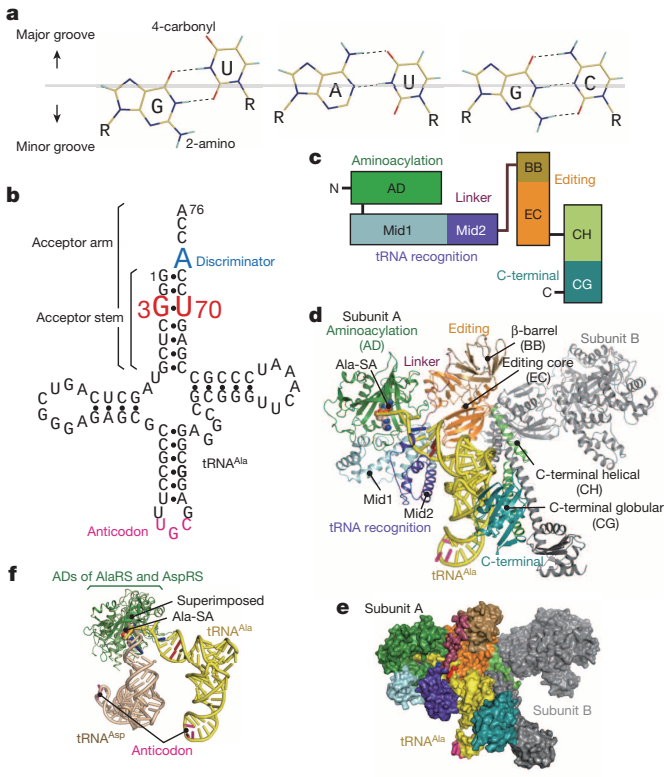


Figure 1 | Structure of AlaRS in complex with tRNA^{Ala}. **a**, G•U, A•U, and G•C pairs. **b**, Cloverleaf model of *A. fulgidus* tRNA^{Ala}. **c**, A diagram of the AlaRS domains/subdomains. **d**, **e**, The *A. fulgidus* AlaRS•tRNA^{Ala}•Ala-SA complex, presented by ribbon (AlaRS) and tube (tRNA^{Ala}) models (**d**) and by surface models (**e**) with the domains/subdomain of AlaRS subunit A coloured as in **c**. **f**, Positions of tRNA^{Ala} and tRNA^{Asp} relative to the class-II aminoacylation domains (superimposed) of AlaRS and AspRS, respectively.

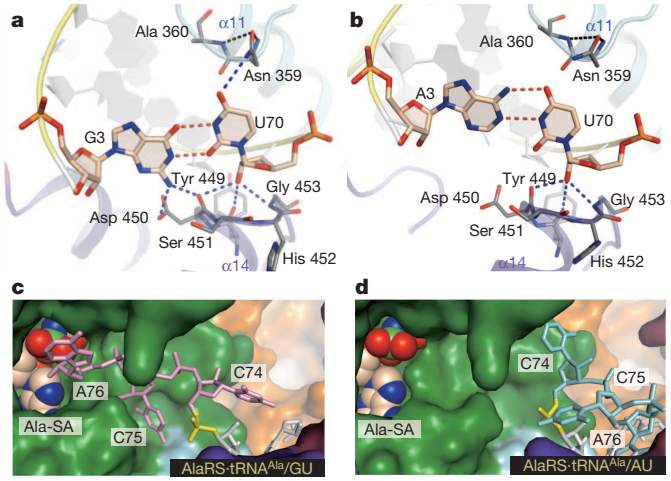


Figure 2 | Major differences between AlaRS•tRNA^{Ala}/GU and AlaRS•tRNA^{Ala}/AU. **a**–**d**, The G3•U70 (**a**) and A3•U70 (**b**) pairs and the CCA regions (**c**, **d**) of tRNA^{Ala}/GU (**a**, **c**) and tRNA^{Ala}/AU (**b**, **d**). The putative hydrogen bonds of G3•U70 and A3•U70 with AlaRS are shown with dashed lines (**a**, **b**). The CCA region of tRNA^{Ala}/GU is in the reactive route (**c**), whereas that of tRNA^{Ala}/AU is in the non-reactive route (**d**). The 5′-phosphate of C74 (P74) is coloured yellow (**c**, **d**).

one tRNA^{Ala} with a K_d value of 0.14 ± 0.3 μM at 20 °C. Between the two subunits, the corresponding subdomains exhibit the same folds. However, the tRNA^{Ala}-bound (subunit A) and tRNA^{Ala}-free (subunit B) subunits are in the closed and open forms, respectively, because the Mid2 subdomain and the C-terminal domain are differently rearranged relative to the other (sub)domains (Fig. 1d, e and Extended Data Fig. 1a). The canonical L-shaped structure of the tRNA is snugly embedded in a large cavity, formed by the closed-form subunit (Fig. 1d, e). The aminoacylation domain of AlaRS adopts the typical class-II aminoacylation-domain fold[32,33], but the position of tRNA^{Ala} relative to the fold is completely different from that conserved in all of the other class-II aaRSs (Fig. 1f and Extended Data Fig. 2). The acceptor–T arm moiety of tRNA^{Ala} interacts with the aminoacylation, tRNA-recognition, and C-terminal domains of AlaRS, whereas the anticodon is free of interactions (Fig. 1d, e). A summary of the AlaRS•tRNA^{Ala} interactions is provided in Extended Data Fig. 3a.

The G3•U70 pair exhibits the wobble geometry: G3 is shifted to the minor-groove side and the two bases are slightly rotated, as compared with the Watson–Crick geometry of A3•U70 (Figs 1a and 2a, b and Extended Data Fig. 4a, b). These geometrical differences are also observed for the model RNA hairpins that recapitulate the tRNA^{Ala} acceptor stem[34,35]. Surprisingly, the two complexes further exhibit a drastic difference in the orientation of the 3′-end CCA region (positions 74–76). The 3′-CCA region of the wild-type tRNA^{Ala}/GU extends towards the aminoacylation catalytic site, so that A76 can reach the bound Ala-SA molecule (Fig. 2c and Extended Data Fig. 5a–c, g, h). However, the entire 3′-CCA region of tRNA^{Ala}/AU folds back into a completely different site (Fig. 2d and Extended Data Fig. 5d–h), thus placing A76 at 20 Å away from the active site. This is an unprecedented mechanism for keeping the bound non-cognate tRNA molecule in a non-reactive state.

The acceptor arm•AlaRS interactions

The minor-groove side of G3•U70 contacts the tip of the α14 helix from the Mid2 subdomain (Fig. 2a). The 2-amino group of G3 may hydrogen bond with the side-chain carboxyl and main-chain carbonyl groups of Asp 450 (Fig. 2a). The 2′-hydroxyl group of U70 possibly hydrogen bonds with the main-chain carbonyl groups of Asp 450 and Tyr 449 and the amino group of Gly 453. The Asp 450 side chain may also interact with the 2′-hydroxyl group of C71 (Extended Data Fig. 4d). Overall, the Tyr-Asp-Ser-His-Gly region (positions 449–453) at the α14 tip fits snugly with the unique minor groove shape of G3•U70 and its adjacent base pairs, G2•C71, C4•G69, and U5•A68 (Fig. 2a and Extended Data Fig. 4d–f). The Tyr-Asp-Ser-His-Gly sequence is conserved as Tyr-(Asp/Glu)-(Ser/Thr)-(His/Tyr)-Gly in other organisms (Extended Data Fig. 4i). The 4-carbonyl oxygen atom of U70 on the major groove side is close to and may hydrogen bond with the side-chain amino group of Asn 359 (Fig. 2a). Overall, the major-groove side from G1•C72 to G3•U70 is located near the helical loop on the N-terminal side of α11, which provides not only Asn 359 but also Gly 361 and Tyr 364 for interactions (Fig. 2a and Extended Data Fig. 4c, d). The base of the discriminator A73 is stacked on both C72 and G1, suggesting hydrogen bonding between A73 and Gly 361 (Extended Data Fig. 4g, h). A similar stacking pattern of A73 with G1•C72 is also observed for the hairpin RNA model[34]. The three interacting residues are almost completely conserved in the AlaRSs from bacteria to humans (Extended Data Fig. 4i).

In the AlaRS•tRNA^{Ala}/AU complex, the acceptor-stem minor groove interacts with the α14 tip, and U in A3•U70 is positioned similarly to that in G3•U70. In contrast, A3 is significantly shifted upward towards the major groove side, and is therefore much farther from Asp 450, as compared with G3 in G3•U70 (Fig. 2a, b). On the major groove side, the 4-carbonyl group of U70 is farther from the Asn 359 side chain, as the uracil base is rotated to pair with the upwardly shifted adenine base. If a G•C pair is substituted for A3•U70, then the 4-amino group of C70 might be more repulsive against Asn 359 than U70. Thus, the local geometry and the interaction manner of A3•U70 (and probably G3•C70) on both the minor and major groove sides are specifically distinct from those of G3•U70.

Geometry propagation from 3•70 to CCA

The local geometrical differences between G3•U70 and A3•U70 on both the minor and major groove sides propagate towards the 3′-CCA region.

©2014 Macmillan Publishers Limited. All rights reserved

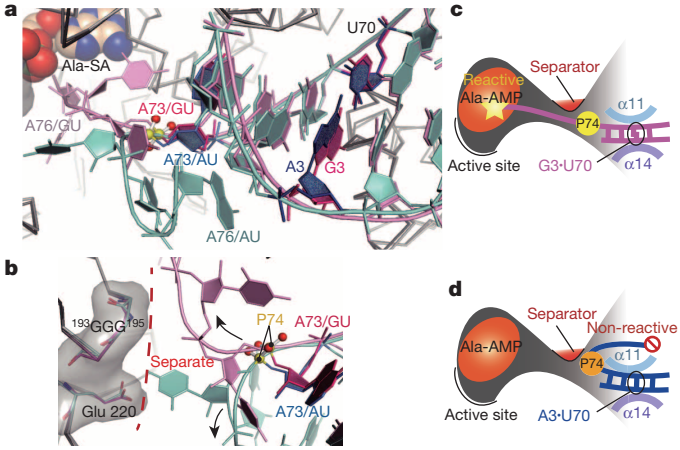


Figure 3 | Watson–Crick pair exclusion. **a**, The acceptor arms of tRNA^{Ala}/GU (pink) and tRNA^{Ala}/AU (cyan), with A73 (red) and G3•U70/A3•U70 (magenta/blue) highlighted. **b**, Route separation for the CCA region, with the turning point at P74, by the separator ^{193}GGG^{195} and Glu 220 (sticks with translucent surface). **c**, **d**, The reactive state of tRNA^{Ala}/GU (**c**) and the non-reactive state of tRNA^{Ala}/AU (**d**).

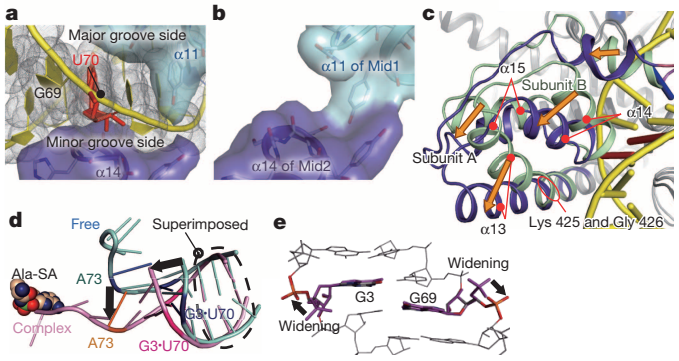


Figure 4 | The tRNA^{Ala} acceptor stem is widened by the clamp. **a**, The tRNA^{Ala} acceptor arm trapped by the clamp (α11 and α14). **b**, The clamp without tRNA^{Ala}. **c**, Mid2 reorientation. Superimposition on Mid1. **d**, The AlaRS-induced widening of the tRNA^{Ala} acceptor stem (pink), as compared to the model RNA (PDB ID: 1IKD) (cyan). **e**, Conformational changes of G3 and G69 relative to the A-form RNAs.

First, because A3•U70 is shifted as compared with G3•U70 and then G2•C71 and G1•C72 are moved together with it (Fig. 3a). These nucleotide movements reach the turning point at the 5′-phosphate group of C74, and thus result in the drastic change in the direction of the polynucleotide backbone (Fig. 3b and Extended Data Fig. 5i). In particular, A76 of tRNA^{Ala}/AU remains near A73, and the two 2′-hydroxyl groups are within hydrogen-bonding distance (< 3.2 Å). Thus, A76 is kept far (about 20 Å) from the catalytic site (Figs 2d and 3a and Extended Data Fig. 5d–f), whereas A76 of tRNA^{Ala}/GU can reach it (Figs 2c and 3a and Extended Data Fig. 5a–c).

The reactive and non-reactive routes for the CCA region are separated by the ^{193}GGG^{195} region and Glu 220, and the CCA region is sterically hindered from shuttling between the two routes (Fig. 3b–d). Consequently, the structure formed by ^{193}GGG^{195} and Glu 220 serves as the route separator. This mechanism of tRNA^{Ala} selection would apply to tRNA^{Ala} variants with Watson–Crick pairs replacing G3•U70 (tRNA^{Ala}/WC). In this context, Asp 174 in *E. coli* AlaRS is spatially near the site corresponding to the ^{193}GGG^{195} region and Glu 220 in *A. fulgidus* AlaRS (Extended Data Fig. 5j, k), and its Gly substitution reportedly causes misaminoacylation of the G3•C70 variant of a model RNA[36].

The geometrical difference between G3•U70 and A3•U70 is smaller than the range of the positional flexibility of the single-stranded CCA region. Consequently, the tRNA selection mechanism directing the CCA region into either the reactive or non-reactive route requires precise positioning of the acceptor stem. The main mechanism is the firm clamping of the major and minor grooves of the acceptor stem by the Mid1 and Mid2 subdomains of AlaRS. First, both the major and minor grooves of the G1•C72, G2•C71, G3•U70 and C4•G69 pairs interact with α11 of Mid1 and α14 of Mid2, respectively (Fig. 4a), whereas more interactions are formed with the 3′-strand (residues 69–73) than the 5′-strand of the acceptor stem (Extended Data Fig. 3a, b). In the tRNA-free subunit B, α13 connecting Mid1 and Mid2 is kinked by ~18°, due to the 3_{10}-like conformation around the conserved Gly 426 near the subdomain boundary (Fig. 4c and Extended Data Fig. 1b, c). In contrast, α13 is straight in the tRNA-bound subunit A, due to the change around Gly 426 from a 3_{10}-like to α-helical conformation (Fig. 4c and Extended Data Fig. 1b, c). Accordingly, α14 and α15 of Mid2 are reoriented upon tRNA binding. This Mid2 reorientation enables Mid1 and Mid2 to snugly clamp the acceptor-stem base pairs (Fig. 4a). In contrast, Mid1 α11 and Mid2 α14 directly contact each other in the tRNA-free subunit B (Fig. 4b).

Concomitantly, the major groove of the acceptor stem is widened by the clamp, as compared with the model RNA hairpin structures[34,35] (Fig. 4d) and the canonical tRNA structures (Extended Data Fig. 6a–f). If the major groove widening did not occur, then the major groove recognition would cause a serious clash of the G68 phosphate group with the editing core subdomain (Extended Data Fig. 6g), and the CCA region could not enter the reactive (or non-reactive) route selectively (Fig. 4d). Specifically, the conformations of the ribose-phosphate backbones of G3 and G69 are both changed from the *gauche*$^-$/*gauche*$^+$ form, typical of the A-form RNA duplex, to the *trans*/*trans* form (Fig. 4e and Extended Data Fig. 7). The *trans*/*trans* conformation has been reported for G3, but not G69, in the NMR structure of the model RNA[34], whereas the crystal structure of a similar model RNA showed the *gauche*$^-$/*gauche*$^+$ form for G3[35]. On the other hand, in the non-reactive AlaRS•tRNA^{Ala}/AU complex, the major-groove of the tRNA^{Ala}/AU acceptor stem is widened in a similar manner to that of tRNA^{Ala}/GU (Extended Data Figs 6h, 7a–c), with almost the same AlaRS conformation as in the wild-type AlaRS•tRNA^{Ala}/GU complex.

In addition to the Mid2 subdomain, the C-terminal domain is remarkably reoriented between the tRNA-bound and tRNA-free subunits (Fig. 4c and Extended Data Fig. 1). The globular subdomain of the C-terminal domain thereby interacts with the outer corner of the L-shaped tRNA molecule. The conserved Gly-rich segment (^{870}KGSGGGR^{876})[28,37] stacks on the G19•C56 tertiary base pair connecting the D and T loops (Extended Data Fig. 3c). Thus, the C-terminal globular subdomain ('backrest') snugly holds the back of tRNA^{Ala}, whereas the tRNA^{Ala} acceptor stem sits on the tRNA recognition domain ('seat'). Furthermore, the long linker (residues 485–500) connecting the Mid2 and β-barrel subdomains ('safety belt') runs over the tRNA^{Ala} acceptor stem, near the minor groove side of G2•C71 (Fig. 1e). These features of the tRNA-embedding mechanism distinguish AlaRS from the other aaRSs, which bind tRNA on their convex surfaces (Fig. 1e and Extended Data Fig. 2).

In the non-reactive conformation of the CCA region of tRNA^{Ala}/AU, A76 is likely to interact with A73 and the linker, which might stabilize the non-reactive orientation. On the other hand, tRNA^{Ala}/AU in the non-reactive conformation hardly seems to dissociate from AlaRS, as the bulky CCA region should prevent the acceptor stem from going through the tunnel between the tRNA recognition domain and the linker (Fig. 2d and Extended Data Fig. 5e, f).

These unique mechanisms for holding the tRNA seem to be required to place the base pair at position 3•70 in the precise recognition site.

Selection of tRNA^{Ala} on the k_{cat} level

By pre-steady state kinetics analyses using a quench flow apparatus, the k_{cat} values of the *A. fulgidus* AlaRS dimer for tRNA^{Ala}/GU and tRNA^{Ala}/AU were determined to be 14.4 ± 0.2 and 0.14 ± 0.02 s^{-1}, respectively. Therefore, the strict tRNA discrimination is achieved by the 100-fold

©2014 Macmillan Publishers Limited. All rights reserved

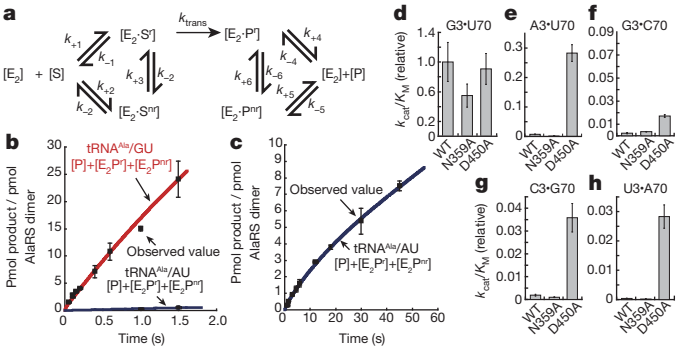


Figure 5 | Alanylation assays of tRNA^{Ala} with AlaRS. **a**, Reaction scheme including the non-reactive state. **b**, **c**, Alanylation of tRNA^{Ala}/GU (red) and tRNA^{Ala}/AU (blue), with curves of $[E_2 \cdot P^r] + [E_2 \cdot P^{nr}] + [P]$ simulated according to **a**. **d**–**h**, The k_{cat}/K_M values of the wild-type, N359A, and D450A AlaRSs for tRNA^{Ala}/GU (**d**) and tRNA^{Ala}/WC (**e**–**h**). Error bars, standard deviations of the mean of three replicates (**b**–**h**).

difference in the k_{cat} value. Next, we performed single turn-over analyses of the alanylation of tRNA^{Ala}/GU and tRNA^{Ala}/AU. Here, k_{chem} is the rate of the overall product formation through two steps, and is almost equal to the rate of the aminoacyl transfer to tRNA, k_{trans}, in the case of *E. coli* AlaRS[38,39]. The k_{chem} (s^{-1}) and K_d (μM) values of the *A. fulgidus* AlaRS dimer for tRNA^{Ala}/AU ($24 \pm 6\ s^{-1}$ and 0.59 ± 0.40 μM, respectively) were only about twofold smaller and larger, respectively, than those of tRNA^{Ala}/GU ($52 \pm 1\ s^{-1}$ and 0.34 ± 0.08 μM, respectively), and thus the differences in the k_{chem} and K_d values between the two tRNAs are much smaller than the two-orders of magnitude difference in the k_{cat} values (s^{-1}).

To interpret these observations, a simplified numerical simulation of the pre-steady state reactions of AlaRS for tRNA^{Ala}/GU and tRNA^{Ala}/AU was performed, on the assumption of a reaction scheme involving the non-reactive state in addition to the reactive state (Fig. 5a). The simulated curves fit the observed data well (Fig. 5b, c), when the reactive state was the major state for tRNA^{Ala}/GU, but the amount of the reactive state was much smaller for tRNA^{Ala}/AU. Here, product release is what limits Ala-tRNA^{Ala}/AU, as it is trapped mostly in the non-reactive state (see Methods). The non-reactive state is most likely to correspond to the present structure of tRNA^{Ala}/AU on AlaRS. Therefore, the strict tRNA selection dependent on k_{cat} can be accomplished by this unprecedented mechanism with the reactive and non-reactive states.

Steady-state kinetics experiments (Fig. 5d–h) also revealed that the discrimination of tRNA^{Ala}/GU from tRNA^{Ala}/WC variants (tRNA^{Ala}/GC, tRNA^{Ala}/AU, tRNA^{Ala}/CG, and tRNA^{Ala}/UA) was mainly due to the two orders of magnitude differences in the k_{cat} values (Extended Data Table 2). These drastic distinctions may correspond to the differences in the probability of adopting the reactive orientation of the CCA region. In addition, we tested k_{cat}-level tRNA^{Ala}/AU discrimination with human AlaRS. Strikingly, human tRNA^{Ala}/AU binds to human AlaRS and competes with tRNA^{Ala}/GU. Moreover, even in the presence of excess amounts of tRNA^{Ala}/AU, the discrimination against human tRNA^{Ala}/AU is ultimately via k_{cat} (Extended Data Fig. 8).

In contrast, the tRNA^{Ala}-binding affinity of AlaRS is essentially the same between tRNA^{Ala}/GU and tRNA^{Ala}/WC, as it is mostly governed by various extensive interactions, such as firmly embedding the tRNA in the huge cavity and snugly clamping the acceptor stem. To examine how the snug clamping contributes to the tRNA selection, we mutated Asn 359 and Asp 450, which directly interact with G3•U70 (Fig. 2a). The k_{cat}/K_M value of the N359A mutant for tRNA^{Ala}/GU was 1.7-fold lower than that of the wild-type AlaRS. By contrast, the D450A mutation affected neither k_{cat} nor K_M (Extended Data Table 2), but remarkably increased the alanylation activity for tRNA^{Ala}/WC (Fig. 5d–h and Extended Data Table 2). These enhanced activities of the D450A mutant are ascribed mostly to an increase in the k_{cat} value (Extended Data Table 2). It is probable that the D450A mutant loosened the clamping of the 3•70 pair, and promoted the misorientation of the CCA region of tRNA^{Ala}/WC towards the reactive route (Extended Data Fig. 5l).

On the other hand, several mismatch base pairs in place of G3•U70 in an alanine suppressor tRNA reportedly cause aminoacylation by AlaRS in *E. coli* cells (G•A, C•A, and C•C are more active and U•U, C•U and A•C are less active), even though their activities are not as high as that of tRNA^{Ala}/GU[40–42]. The effects of these mismatch base pairs may be analogous to those of the D450A mutation, which incorrectly guides the CCA region into the reactive route.

This unprecedented mechanism now provides an answer to the long-standing question of how a small number of nucleotides about 30 Å away from the catalytic site can predominantly determine the tRNA identity, particularly on the k_{cat} level.

METHODS SUMMARY

The wild-type and mutant *A. fulgidus* AlaRSs were prepared by the method reported for AlaRS fragment preparation[22,28]. AlaRS was co-crystallized with the wild-type or variant *A. fulgidus* tRNA^{Ala} transcript prepared *in vitro* with T7 RNA polymerase. The X-ray data were collected at the Photon Factory and SPring-8. The initial phases were determined by the molecular replacement method, using the coordinates of the N- and C-terminal fragments of *A. fulgidus* AlaRS (PDB IDs: 2ZTG and 2ZVF, respectively), and *Saccharomyces cerevisiae* tRNA^{Asp} (PDB ID: 1IL2)[32] as search models. The refinement converged to R and R_{free} factors of 25.0% and 28.9%, respectively. Aminoacylation activities were measured by scintillation counting of the radioactive product from L-[^{3}H]alanine.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 February; accepted 6 May 2014.
Published online 11 June 2014.

1. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26,** 5017–5035 (1998).
2. Beuning, P. J. & Musier-Forsyth, K. Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers* **52,** 1–28 (1999).
3. Pütz, J., Puglisi, J. D., Florentz, C. & Giegé, R. Additive, cooperative and anti-cooperative effects between identity nucleotides of a tRNA. *EMBO J.* **12,** 2949–2957 (1993).
4. Weygand-Durasević, I., Rogers, M. J. & Söll, D. Connecting anticodon recognition with the active site of *Escherichia coli* glutaminyl-tRNA synthetase. *J. Mol. Biol.* **240,** 111–118 (1994).
5. Rogers, M. J., Adachi, T., Inokuchi, H. & Söll, D. Functional communication in the recognition of tRNA by *Escherichia coli* glutaminyl-tRNA synthetase. *Proc. Natl Acad. Sci. USA* **91,** 291–295 (1994).
6. Uter, N. T. & Perona, J. J. Long-range intramolecular signaling in a tRNA synthetase complex revealed by pre-steady-state kinetics. *Proc. Natl Acad. Sci. USA* **101,** 14396–14401 (2004).
7. Hou, Y. M. & Schimmel, P. A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333,** 140–145 (1988).
8. McClain, W. H. & Foss, K. Changing the identity of a tRNA by introducing a G-U wobble pair near the 3′ acceptor end. *Science* **240,** 793–796 (1988).
9. de Duve, C. Transfer RNAs: the second genetic code. *Nature* **333,** 117–118 (1988).
10. Park, S. J., Hou, Y. M. & Schimmel, P. A single base pair affects binding and catalytic parameters in the molecular recognition of a transfer RNA. *Biochemistry* **28,** 2740–2746 (1989).
11. Hou, Y. M. & Schimmel, P. Evidence that a major determinant for the identity of a transfer RNA is conserved in evolution. *Biochemistry* **28,** 6800–6804 (1989).
12. Shi, J. P., Francklyn, C., Hill, K. & Schimmel, P. A nucleotide that enhances the charging of RNA minihelix sequence variants with alanine. *Biochemistry* **29,** 3621–3626 (1990).
13. Beuning, P. J., Gulotta, M. & Musier-Forsyth, K. Atomic group "mutagenesis" reveals major groove fine interactions of a tRNA synthetase with an RNA helix. *J. Am. Chem. Soc.* **119,** 8397–8402 (1997).
14. Musier-Forsyth, K. & Schimmel, P. Functional contacts of a transfer RNA synthetase with 2′-hydroxyl groups in the RNA minor groove. *Nature* **357,** 513–515 (1992).
15. Beuning, P. J. *et al.* Efficient aminoacylation of the tRNA^{Ala} acceptor stem: dependence on the 2:71 base pair. *RNA* **8,** 659–670 (2002).
16. Francklyn, C., Shi, J. P. & Schimmel, P. Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science* **255,** 1121–1125 (1992).
17. Francklyn, C. & Schimmel, P. Aminoacylation of RNA minihelices with alanine. *Nature* **337,** 478–481 (1989).
18. Varani, G. & McClain, W. H. The G•U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* **1,** 18–23 (2000).
19. Musier-Forsyth, K. *et al.* Specificity for aminoacylation of an RNA helix: an unpaired, exocyclic amino group in the minor groove. *Science* **253,** 784–786 (1991).

©2014 Macmillan Publishers Limited. All rights reserved

20. Chang, K. Y., Varani, G., Bhattacharya, S., Choi, H. & McClain, W. H. Correlation of deformability at a tRNA recognition site and aminoacylation specificity. *Proc. Natl Acad. Sci. USA* **96,** 11764–11769 (1999).
21. Ueda, H. *et al.* X-ray crystallographic conformational study of 5′-O-[N-(L-alanyl)-sulfamoyl]adenosine, a substrate analogue for alanyl-tRNA synthetase. *Biochim. Biophys. Acta* **1080,** 126–134 (1991).
22. Fukunaga, R. & Yokoyama, S. Crystallization and preliminary X-ray crystallographic study of alanyl-tRNA synthetase from the archaeon *Archaeoglobus fulgidus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63,** 224–228 (2007).
23. Dignam, J. D. *et al.* Allosteric interaction of nucleotides and $tRNA^{ala}$ with *E. coli* alanyl-tRNA synthetase. *Biochemistry* **50,** 9886–9900 (2011).
24. Jasin, M., Regan, L. & Schimmel, P. Modular arrangement of functional domains along the sequence of an aminoacyl tRNA synthetase. *Nature* **306,** 441–447 (1983).
25. Beebe, K., Ribas De Pouplana, L. & Schimmel, P. Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. *EMBO J.* **22,** 668–675 (2003).
26. Swairjo, M. A. *et al.* Alanyl-tRNA synthetase crystal structure and design for acceptor-stem recognition. *Mol. Cell* **13,** 829–841 (2004).
27. Guo, M. *et al.* Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. *Nature* **462,** 808–812 (2009).
28. Naganuma, M., Sekine, S., Fukunaga, R. & Yokoyama, S. Unique protein architecture of alanyl-tRNA synthetase for aminoacylation, editing, and dimerization. *Proc. Natl Acad. Sci. USA* **106,** 8489–8494 (2009).
29. Sokabe, M. *et al.* The structure of alanyl-tRNA synthetase with editing domain. *Proc. Natl Acad. Sci. USA* **106,** 11028–11033 (2009).
30. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347,** 203–206 (1990).
31. Cusack, S. Eleven down and nine to go. *Nature Struct. Biol.* **2,** 824–831 (1995).
32. Ruff, M. *et al.* Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science* **252,** 1682–1689 (1991).
33. Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. The 2.9 Å crystal structure of T. *thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science* **263,** 1404–1410 (1994).
34. Ramos, A. & Varani, G. Structure of the acceptor stem of *Escherichia coli* $tRNA^{Ala}$: role of the G3·U70 base pair in synthetase recognition. *Nucleic Acids Res.* **25,** 2083–2090 (1997).
35. Mueller, U., Schübel, H., Sprinzl, M. & Heinemann, U. Crystal structure of acceptor stem of $tRNA^{Ala}$ from *Escherichia coli* shows unique G•U wobble base pair at 1.16 Å resolution. *RNA* **5,** 670–677 (1999).
36. Miller, W. T., Hou, Y. M. & Schimmel, P. Mutant aminoacyl-tRNA synthetase that compensates for a mutation in the major identity determinant of its tRNA. *Biochemistry* **30,** 2635–2641 (1991).
37. Guo, M. *et al.* The C-Ala domain brings together editing and aminoacylation functions on one tRNA. *Science* **325,** 744–747 (2009).
38. Zhang, C. M., Perona, J. J., Ryu, K., Francklyn, C. & Hou, Y. M. Distinct kinetic mechanisms of the two classes of Aminoacyl-tRNA synthetases. *J. Mol. Biol.* **361,** 300–311 (2006).
39. Guth, E., Connolly, S. H., Bovee, M. & Francklyn, C. S. A substrate-assisted concerted mechanism for aminoacylation by a class II aminoacyl-tRNA synthetase. *Biochemistry* **44,** 3785–3794 (2005).
40. McClain, W. H., Chen, Y. M., Foss, K. & Schneider, J. Association of transfer RNA acceptor identity with a helical irregularity. *Science* **242,** 1681–1684 (1988).
41. Gabriel, K., Schneider, J. & McClain, W. H. Functional evidence for indirect recognition of G•U in $tRNA^{Ala}$ by alanyl-tRNA synthetase. *Science* **271,** 195–197 (1996).
42. McClain, W. H., Jou, Y. Y., Bhattacharya, S., Gabriel, K. & Schneider, J. The reliability of *in vivo* structure-function analysis of tRNA aminoacylation. *J. Mol. Biol.* **290,** 391–409 (1999).

Acknowledgements We are grateful to the staff of the beam line AR-NW12A at the Photon Factory and the beam lines BL32XU and BL41XU at SPring-8 for their assistance during data collection. This work was supported by the Targeted Proteins Research Program, and the Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. This work was also supported by a Grant-in-Aid for Scientific Research (No. 20247008) from the Japan Society for the Promotion of Science and MEXT, and the Global COE Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms) from MEXT. Research support was also provided by National Institutes of Health grants GM015539 and GM023562 to P.S. and NS085092 to X.-L.Y.

Author Contributions M.N., S.S. and S.Y. designed the research. M.N. and S.S. performed the structural analysis. M.N., Y.E.C., M.G., X.-L.Y., H.G., and Y.-M.H. performed the enzyme kinetics analyses. M.N., S.S., P.S., and S.Y. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The coordinates and structure factors for AlaRS complexed with its cognate $tRNA^{Ala}$ with G3•U70 and a variant with A3•U70 have been deposited in the Protein Data Bank (PDB) under accession codes 3WQY and 3WQZ, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S. Y. (yokoyama@riken.jp).

©2014 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature13423

Structure of a modular polyketide synthase

Somnath Dutta[1]*, Jonathan R. Whicher[1,2]*, Douglas A. Hansen[1,3], Wendi A. Hale[4], Joseph A. Chemler[1], Grady R. Congdon[1], Alison R. H. Narayan[1], Kristina Håkansson[4], David H. Sherman[1,3,4,5], Janet L. Smith[1,6] & Georgios Skiniotis[1,6]

Polyketide natural products constitute a broad class of compounds with diverse structural features and biological activities. Their biosynthetic machinery, represented by type I polyketide synthases (PKSs), has an architecture in which successive modules catalyse two-carbon linear extensions and keto-group processing reactions on intermediates covalently tethered to carrier domains. Here we used electron cryo-microscopy to determine sub-nanometre-resolution three-dimensional reconstructions of a full-length PKS module from the bacterium *Streptomyces venezuelae* that revealed an unexpectedly different architecture compared to the homologous dimeric mammalian fatty acid synthase. A single reaction chamber provides access to all catalytic sites for the intramodule carrier domain. In contrast, the carrier from the preceding module uses a separate entrance outside the reaction chamber to deliver the upstream polyketide intermediate for subsequent extension and modification. This study reveals for the first time, to our knowledge, the structural basis for both intramodule and intermodule substrate transfer in polyketide synthases, and establishes a new model for molecular dissection of these multifunctional enzyme systems.

Type I modular polyketide synthases are enzyme assembly lines for the synthesis of chemically diverse polyketide natural products that form the basis for nearly one-third of pharmaceuticals[1]. Accordingly, it is essential to dissect the architecture of the PKS module and to elucidate the molecular basis for selectivity and catalysis to enable rational bioengineering efforts for the creation of high-value chemicals and novel drugs.

Type I PKS modules act successively in polyketide chain elongation, processing and termination (Fig. 1)[2]. Each module contains acyl carrier protein (ACP), ketosynthase (KS) and acyltransferase (AT) domains that extend the linear sequence of an intermediate by two carbon atoms. The AT loads the ACP with a building block from a specific acyl-coenzyme A (acyl-CoA), and the KS catalyses carbon–carbon bond formation between the intermediate from the upstream module and the acyl-ACP. In addition, modules may also contain domains that successively modify the β-keto group to a hydroxyl (ketoreductase (KR)), a double bond (dehydratase (DH)), or a single bond (enoylreductase (ER)). The ACP employs a phosphopantetheine (Ppant) arm and thioester bond to tether polyketide intermediates and building blocks, and transfers them to respective catalytic domains for loading, extension and keto-group processing or modification. Through direct fusion or a non-covalent docking interaction, the ACP also transfers the fully processed intermediate to the KS in the subsequent module for further extension or to the thioesterase (TE) in the final module for release through hydrolysis or cyclization.

Modular PKSs are thought to share a common ancestor with the mammalian fatty acid synthase (FAS)[3]. However, to achieve perfect fidelity in product formation, the sequential modular PKSs have evolved to be highly ordered in structure and function. The FAS on the other hand acts iteratively within a single multifunctional protein with no inherent ability to deviate from its role to construct saturated fatty acid chains. Whereas the FAS accesses a single ACP, the PKS must ensure that the many ACPs within a multimodule assembly line deliver their respective polyketide intermediates with absolute fidelity to the appropriate modules and catalytic domains. Furthermore, PKSs have evolved through insertion or deletion of catalytic domains to vary extensively in the series of reactions performed within a single module[2].

Currently, no high-resolution structure is available for any PKS module, although crystal structures have been reported for various excised PKS catalytic domains, KS–AT di-domains and docking domains[4–21]. Despite the important insights from these studies, the working model for the PKS module is based on a crystal structure of the porcine FAS[22], which has several limitations due to differences in the oligomeric state and structure of individual catalytic domains as well as substantial differences in linker regions. Thus, obtaining high-resolution information on the overall structure and organization of a complete PKS module is central for understanding polyketide extension, processing and diversification.

Electron cryo-microscopy of PikAIII

To investigate the structure of a full-length type I PKS module, we applied single-particle electron cryo-microscopy (cryo-EM) to visualize pikromycin PKS module 5 (PikAIII) from *Streptomyces venezuelae* (Fig. 1)[23]. The mono-module polypeptide, PikAIII, bearing the KS_5–AT_5–KR_5–ACP_5 domain architecture (Fig. 1), can be produced in pure, stable form (Extended Data Fig. 1a) and characterized by established biochemical assays[24,25].

Cryo-EM three-dimensional reconstructions of PikAIII were obtained in different physiological states. In conditions in which more than one PikAIII conformation was observed, we followed a multiple-reference classification approach[26,27] to separate cryo-EM projections and calculate independent three-dimensional maps of each conformer. On the basis of micrograph quality control and inclusion of several thousands of cryo-EM images, we obtained final reconstructions with indicated resolutions of 7.3–9.5 Å (Extended Data Figs 1–4). The three-dimensional maps revealed secondary structures within PikAIII and allowed precise rigid body fitting of the homologous structures of the KS, AT, KR and ACP domains from 6-deoxyerythronolide B synthase (DEBS)[7,11,20], leading to highly accurate pseudo-atomic-resolution structures of PikAIII.

[1]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, USA. [2]Chemical Biology Graduate Program, University of Michigan, Ann Arbor, Michigan 48109, USA. [3]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [4]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [5]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan 48109, USA. [6]Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

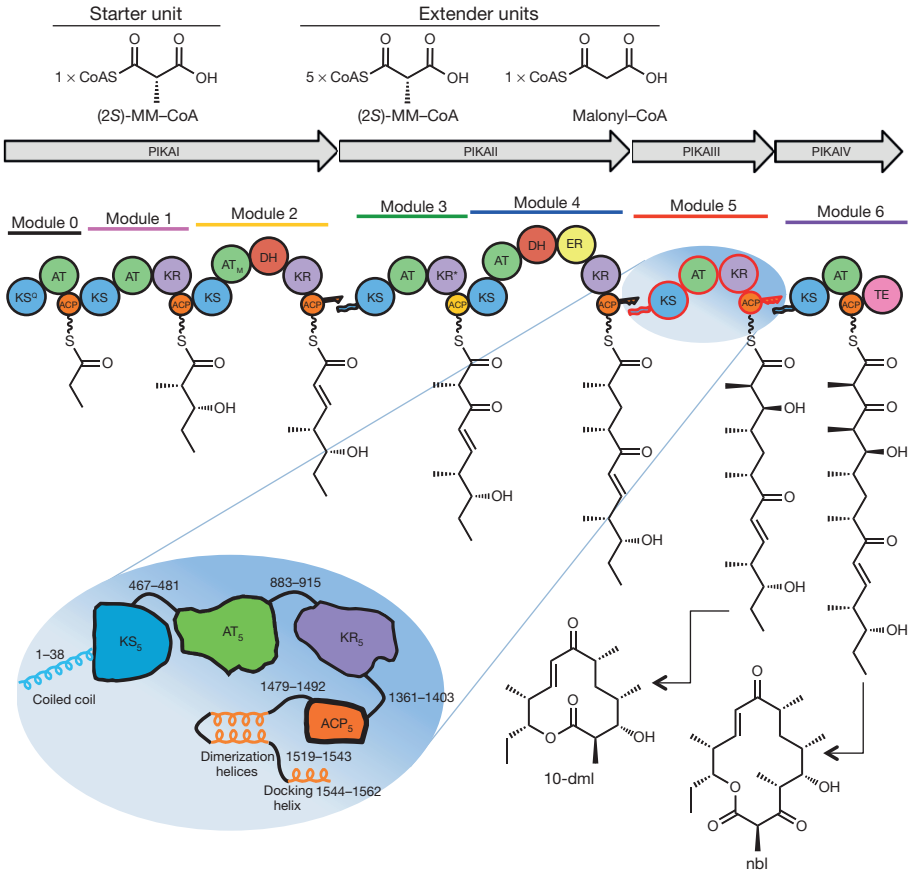


Figure 1 | Modular polyketide synthase for pikromycin. The six modules of the pikromycin PKS, comprising PikAI–IV polypeptides, sequentially elongate and modify a polyketide intermediate. A polyketide product, either 10-dml from module 5 or nbl from module 6, is off-loaded by the TE domain of the final module, PikAIV. Modules are differently coloured. Circles represent protein domains (KS, AT, DH, ER, KR and ACP; KS^Q is a decarboxylase; KR* is inactive), and docking domains are shown as jagged ends. PikAIII schematic: the 1,562-amino-acid PikAIII polypeptide, selected for this study, is shown with functional domains in contrasting colours, used throughout, and linker peptides identified by residue ranges. The N- and C-terminal docking domains are shown as helices, as are the post-ACP dimerization helices.

Thus, here we describe the first structural snapshots—to our knowledge—of a full-length PKS module in different functional states.

Overall structure of the PKS module

The cryo-EM structure of the 328 kilodalton (kDa) dimeric PikAIII with the intramodule ACP_5 bearing a Ppant arm (holo-PikAIII) (Extended Data Figs 1b, c, 5a–d) reveals an arch-shaped symmetric dimer, with the homodimeric KS_5 at the dome supported on either side through a post formed by an AT_5 followed by KR_5 at the base (Fig. 2a–c). The ACP_5 domains, which fit unambiguously into their corresponding densities (Fig. 2d), are found to populate two distinct positions in a ~1:1 ratio, one near KR_5, and one near AT_5 (Fig. 2 and Extended Data Fig. 4). In both conformers the remaining domains display identical configurations. Bottom-up liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC/FT-ICR MS) established that the Ppant arm was present at a high level (Extended Data Fig. 5a–d), suggesting that the distinct ACP_5 locations are not attributable to the presence or absence of this post-translational modification. In addition, multiple-reference-supervised classification tests could not detect a statistically significant fraction of holo-PikAIII dimers with the two ACP_5 domains in non-equivalent positions, that is, one near the AT_5 and the other near the KR_5, providing evidence for concerted ACP_5 movement.

The KS_5–AT_5 configuration in the complete PikAIII module resembles an earlier proposed PKS model[28] but is strikingly different from the configuration of the excised KS–AT di-domains[6,7,21] (Extended Data Fig. 6a) and the mammalian FAS[22] (Fig. 2e). In full-length PikAIII, each AT is rotated by ~120° relative to its position in the FAS and excised KS–ATs, and forms an extensive interface with KS_5 (Extended Data Fig. 7a). In this arrangement, AT_5 and the following KR_5 domain reside below KS_5, creating a single ACP_5 reaction chamber in the centre of the PikAIII dimer, with the AT and KR active sites facing the chamber. The characteristic long and narrow shape of the PikAIII reaction chamber is probably crucial for excluding ACPs of other modules, thus providing a mechanism to maintain biosynthetic fidelity.

The three-dimensional maps of PikAIII reveal weak density for the post-ACP_5 dimerization helices (Extended Data Fig. 6b, c) and no density for the interdomain connecting loops, either because of insufficient resolution, disorder, or a combination of both. Nevertheless, linker lengths and distance limitations indicate that each AT_5–KR_5 post includes domains of the same chain, while each AT interfaces with the KS of the opposite polypeptide (Fig. 1 and Extended Data Fig. 6d). The linker from KR_5 to ACP_5 (43 amino acids) is long enough to facilitate ACP interaction with the KS of either subunit, suggesting that other types of steric constraints may limit each ACP to interact with the KS domain of the opposite monomer, as reported for a DEBS module[29,30]. Only the KS_5–AT_5 linker (15 amino acids) is too short, requiring an additional eight amino acids to span the shortest distance between a KS_5 carboxy terminus and an AT_5 amino terminus (70 Å). This observation suggests that some amino acids at the AT N terminus of excised KS–ATs may be part of the KS–AT linker in the full module. While the resolution of the EM maps does not allow us to confirm this hypothesis, several lines of evidence indicate that the KS_5–AT_5 interaction is associated with structural rearrangements at the domain interface. First, this is the only region of domain clashes between the crystal structures rigidly docked in the

©2014 Macmillan Publishers Limited. All rights reserved

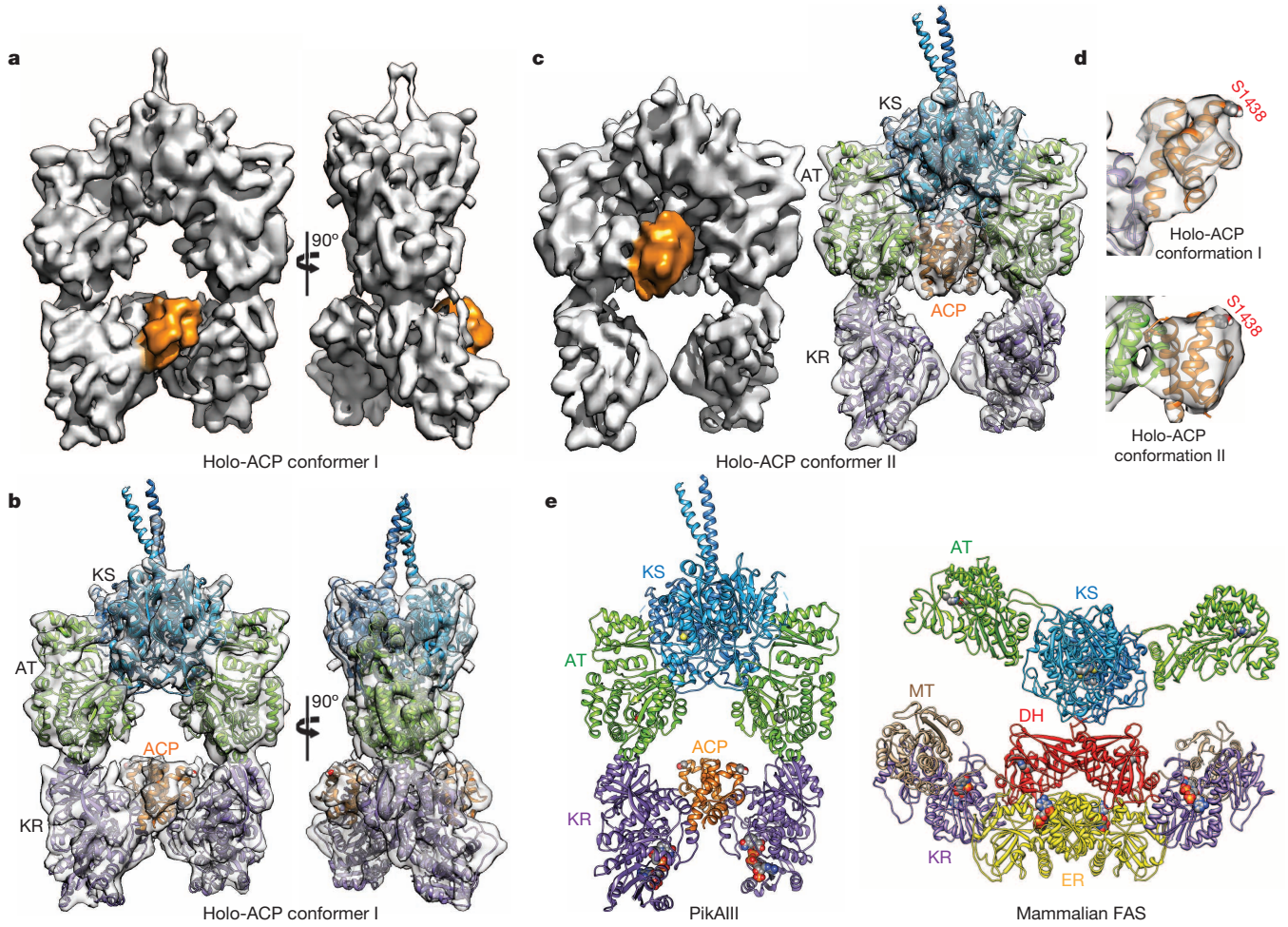


Figure 2 | Cryo-EM structures of holo-PikAIII. **a**, **b**, Solid rendering (**a**) and transparent representations with modelled structures (**b**) of the cryo-EM map of holo-PikAIII conformer I. This conformation, in which the ACP_5 (orange) is near KR_5, is observed in 57% of particle projections. **c**, Solid rendering (left) and transparent representations with modelled structures (right) of the cryo-EM map of holo-PikAIII conformer II. This conformation, in which ACP_5 (orange) is near AT_5, is observed in 43% of particle projections. **d**, Fitting of the DEBS ACP structure[20] in the corresponding density of both holo-PikAIII conformers reveals that Ser 1438 is directed away from any active site. **e**, Comparison of holo-PikAIII full module and mammalian FAS[22]. The KS is blue, AT is green, ACP is orange (disordered in FAS), DH is yellow, and ER is red. Active sites are highlighted with spheres.

three-dimensional maps, and structural rearrangements must occur to alleviate these clashes (Extended Data Fig. 7a, b). Furthermore, the position of AT_5 relative to KS_5 changes according to the substrate condition of PikAIII, indicating a dynamic KS_5–AT_5 interface that adapts to different states during the PikAIII catalytic cycle (see accompanying paper[31]).

Mutagenesis of PikAIII domain interfaces

To probe the KS_5–AT_5 interaction and the associated conformational changes, we generated KS_5 and AT_5 mutants in PikAIII–TE (PikAIII fused to the TE domain of PikAIV)[24,25,32] and assessed their ability to generate 10-deoxymethynolide (10-dml) (Extended Data Fig. 8a, b), which is a natural macrolactone product of the Pik PKS[24,25]. PikAIII AT_5 Lys 490 and KS_5 Asp 352 are near one another within the KS_5–AT_5 clash zone whereas Arg 525 is outside the clash zone (Extended Data Fig. 7b). Mutagenesis at these sites highlighted a critical role for Arg 525, as also observed in the DEBS system[33]. The results suggest that conformational changes within AT_5 accompany its interaction with KS_5, whereby Arg 525 is probably crucial to the domain interface while Lys 490 is outside the interaction zone. Furthermore, these AT_5 amino acids are near the KS_5 active-site entrance in full-length PikAIII, but far away in the excised KS–AT structures, providing further validation of the cryo-EM structure.

Linker distance limitations suggest that each AT interacts with the KS of the opposite chain but with the KR of the same monomer (Extended Data Fig. 6d). The AT–KR conformation is incompatible with anchoring of the post-AT linker (residues 891–903) to the side of a KS monomer (Extended Data Fig. 6a), as observed in crystal structures of excised KS–AT di-domains and mammalian FAS[6,7,21,22], consistent with the lack of any corresponding density in the cryo-EM maps. To confirm this finding, we mutated conserved Phe 897 and Trp 903, which make extensive KS contacts in the excised KS–AT crystal structures (Extended Data Fig. 8b). PikAIII W903A and F897A exhibited 7% and 130% of wild-type activity, respectively, suggesting that Phe 897 does not tether the post-AT_5 linker in the position found in the crystal structures, and that Trp 903 may have some other important functional role. We probed the AT_5–KR_5 interface (Extended Data Fig. 7c) with three single-site substitutions, E766R and E768R in AT_5 and G929E in KR_5. The AT_5 mutations reduced activity (77%, E766R; 45%, E768R) compared to wild type, while the KR_5 mutation had no effect, indicating that the observed AT_5–KR_5 interface may not be crucial for PikAIII function (Extended Data Fig. 8b).

KS_5 active-site entrance for PikAII ACP_4

To examine how the upstream ACP_4 domain (PikAII module 4; Fig. 1) docks and transfers the pentaketide intermediate to PikAIII KS_5, we created a PikAIII lacking ACP_5 to allow visualization of the upstream ACP_4 without the added complexity of the intramodule ACP. Unexpectedly, we found that a PikAIII lacking ACP_5 and its corresponding

©2014 Macmillan Publishers Limited. All rights reserved

dimerization helices (PikAIII(Δ1403–1562)) was monomeric (Extended Data Fig. 9a), while a PikAIII with an internal ACP deletion (ΔACP_5) but including the post-ACP_5 sequences (PikAIII(ΔACP_5)) was dimeric (Extended Data Fig. 9a). This finding suggests that PikAIII dimerization by KS_5 is relatively weak and formation of a stable dimer requires the additional avidity provided by the post-ACP dimerization helices. The structure of PikAIII(ΔACP_5) at 7.8 Å (Extended Data Figs 4, 9b) revealed the same overall architecture of the PikAIII module as shown for holo-PikAIII, but with KR_5 domains rotated by 165° about each leg of the PikAIII arch (Extended Data Fig. 7d). Thus, the orientation of the KR_5 domains with respect to the AT_5 domains appears to be strongly influenced by the presence of the intramodule ACPs.

We next examined a PikAIII(ΔACP_5) variant bearing ACP_4 fused with a flexible linker to the N-terminal docking domain of KS_5. Fusion of the ACP_4 was necessary owing to the weak affinity of docking domains (dissociation constant (K_d) of 5 μM)[5]. The KS_5 active-site Cys 209 was substituted with an alanine (ACP_4–PikAIII(C209A/ΔACP_5)) to prevent intermediate transfer from a pentaketide-loaded ACP_4 (Extended Data Fig. 5e, f). Bottom-up LC/FT-ICR MS confirmed pentaketide loading onto the ACP_4–PikAIII(C209A/ΔACP_5) fusion protein (Extended Data Figs 5g–l, 9c, d). The structure of pentaketide–ACP_4–PikAIII(C209A/ΔACP_5) at 8.6 Å revealed ACP_4 bound on the top of KS, completely outside the intramodule ACP_5 chamber (Fig. 3a). The structure of the DEBS ACP[20] was accurately docked in its corresponding density near the entrance of the KS_5 active-site channel, revealing that ACP_4 Ser 3605 is near the KS_5 catalytic Cys 209 (Fig. 3b).

To probe the interaction between upstream ACP and downstream KS, we mutagenized the interaction region of PikAIV (module 6) KS_6. The well-established PikAIII/PikAIV[24,34] assay generates two products, which are formed by transfer of the β-hydroxyhexaketide from PikAIII ACP_5 to the PikAIV KS_6 (following chain extension to the heptaketide and cyclization to generate narbonolide (nbl)) or directly to the PikAIV TE (to generate 10-dml) (Fig. 1 and Extended Data Fig. 8c). Therefore, formation of nbl requires a functional ACP_5–KS_6 interaction while formation of 10-dml does not[34]. PikAIV KS_6 residues Arg 147 and Arg 320 are analogous to PikAIII KS_5 basic residues at the ACP_4–KS_5 interface in the cryo-EM reconstructions. Substitutions R147E, R320E, R320H and R147E/R320H reduced nbl formation to 83%, 60%, 80% and 44%, respectively, compared to wild-type PikAIV, but did not affect 10-dml formation (110%, 107%, 115% and 127%, respectively) (Extended Data Fig. 8d).

Interestingly, no density for the upstream ACP_4 was observed in the cryo-EM map of unloaded holo-ACP_4–PikAIII(C209A/ΔACP_5) (Extended Data Fig. 9e, f). These findings provide strong evidence that the substrate intermediate loaded on upstream ACP_4 is a primary driver of docking at the KS_5 active-site entrance, in addition to protein–protein interactions as identified in the Pik and DEBS systems[5,34–36].

KS_5 active-site entrance for PikAIII ACP_5

Next, we probed how the intramodule ACP_5 engages the KS_5 active site when it carries the ACP_5-bound methylmalonyl (MM) extender unit for elongation. To prepare MM-loaded ACP_5, we incubated holo-PikAIII with MM–CoA and confirmed the presence of MM–ACP_5 and MM–AT_5 (Extended Data Fig. 5m–s)[37] by bottom-up LC/FT-ICR MS. The cryo-EM structure of MM–PikAIII at 7.3 Å resolution (Fig. 4a) showed the same overall PikAIII conformation. However, MM–ACP_5 interacts with a loop (residues 83–92) at the bottom of the KS_5 within the ACP_5 chamber, further indicating that ACPs localize according to the tethered substrate. Despite the overall high resolution of the MM–PikAIII cryo-EM map, the ACP does not fit its corresponding density with the remarkable accuracy we found in the other states, suggesting considerable flexibility. The lack of ACP positional rigidity is expected, considering the flexibility of loops at this site in the KS–AT crystal structures[6,7,21]. Nevertheless, the conformation that provides the highest cross-correlation between the ACP_5 model and the EM density is consistent with previous ACP mutagenesis data[6,35,36] and places the ACP_5 Ser 1438 below the KS_5 active site at a distance of 25 Å (Extended Data Fig. 7e). To confirm this interaction, we produced single-site loop substitutions, which abolished macrolactone formation (R91A) or decreased it (D87A, 30% reduction) (Extended Data Fig. 8b).

The ACP_5 location in MM–PikAIII is far from the previously established KS_5 active-site entrance, suggesting the existence of a second entrance at the bottom of KS_5. Indeed, we find that the KS domains of modular PKSs have a previously unrecognized active-site entrance, which faces the central ACP_5 chamber of PikAIII (Fig. 4b) and is surrounded by a loop (163–174 in PikAIII) with low sequence conservation and high flexibility in the KS–AT crystal structures[6,7,21]. To confirm the functional role of the channel, we sought to obstruct it by substituting tryptophan for amino acids surrounding the channel entrance. Consistent with our hypothesis, the single substitutions S164W, N241W and D243W showed 4%, 12% and 6% of wild-type activity, respectively (Extended Data Fig. 8b). Taken together, the MM–PikAIII structure, mutagenesis results and sequence alignments across many type I modular PKSs reveal a new active-site entrance that faces into the KS–AT–KR catalytic chamber. The entrance is used by the intramodule ACP to transfer substrates to

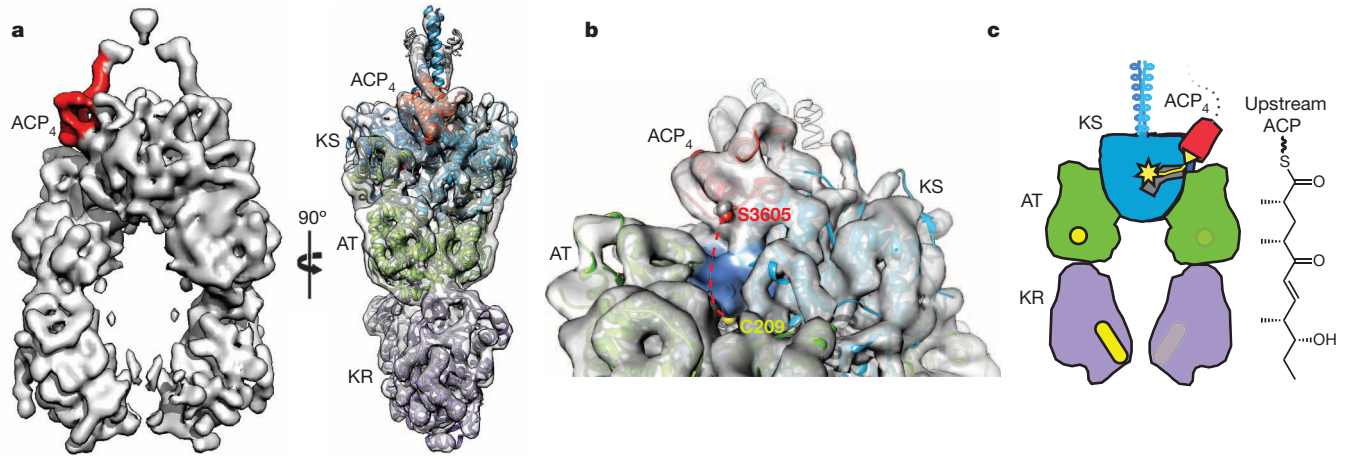


Figure 3 | Interaction of upstream ACP with the PKS module. a, Orthogonal views of solid rendering (left) and transparent representation with modelled structures (right) of the cryo-EM map of pentaketide–ACP_4–PikAIII(C209A/ΔACP_5) at 8.6 Å resolution. **b**, The position of Ser 3605 proximal to the KS active-site entrance. Ser 3605 (red sphere) and Cys 209 (yellow sphere) are 28 Å apart (dashed red line). Loops 1 and 2 of ACP_4 (residues 3588–3606 and 3624–3634) contact two helices (residues 284–293 and 316–322) and a loop (residues 140–150) of KS_5. **c**, Cartoon representation of pentaketide–ACP_4–PikAIII(C209A/ΔACP_5). The upstream ACP (red with yellow serine) carrying the pentaketide intermediate (yellow line) docks to the side entrance of the downstream KS (blue with yellow active site).

©2014 Macmillan Publishers Limited. All rights reserved

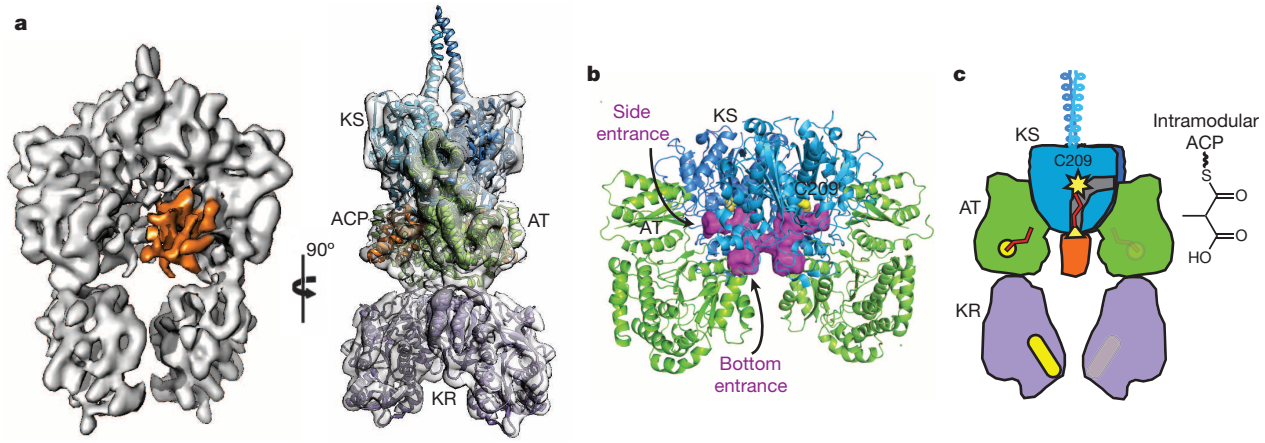


Figure 4 | Interaction of intramodule ACP_5 with KS_5 in MM–PikAIII. **a**, Orthogonal views of solid rendering (left) and transparent representation with modelled structures (right) of the MM–PikAIII cryo-EM map at 7.3 Å resolution. The ACP (orange) has shifted ~20 Å relative to its position in holo-PikAIII conformer II. **b**, KS active-site channels. Internal cavity analysis (purple surface) depicts channels to the active site from both the side and bottom entrances. **c**, Cartoon representation of MM–PikAIII. AT (green, yellow active site) loading of the MM building block (red) onto the intramodule ACP (orange, yellow serine) positions the carrier domain at the bottom entrance of KS (blue, yellow active site) for decarboxylative condensation, remote from the KR domain (purple, yellow active site).

and from the KS. In contrast, FAS I and II, as well as type II and III PKSs, lack this second entrance[22,38–41]. Instead, the bottom channel is closed by well-ordered loops and the channel harbouring the KS catalytic cysteine has only one entrance.

Discussion

The PikAIII structure, strikingly different from the more open architecture of mammalian FAS with its two ACP chambers[22] (Fig. 2e), reveals nature's remarkable design of a dimeric type I PKS module with a single chamber that sequesters the active-site entrances so that they are readily accessible to the intramodule ACP and unlikely to encounter incorrect substrates delivered by other near-neighbour ACPs. This architecture facilitates efficient substrate channelling by providing separate entrances to the KS active site for the ACP carrying an intermediate from the upstream module and for the intramodule ACP that carries the next extender unit and, subsequently, the elongated intermediate (Figs 3c, 4c). The newly identified intramodule ACP entrance on the KS is almost certainly inaccessible to the upstream ACP, which is docked to the KS coiled coil. It is less certain whether the intramodule ACP can reach the KS entrance for the upstream ACP, as the linker peptides were not visible in our cryo-EM maps. In addition, the intramodule KS entrance appears to be a unique feature of bacterial type I modular PKSs as it was not observed in iterative type II PKS, ACP-less type III PKS, mammalian FAS, yeast FAS and bacterial type II FAS structures[22,38–40,42]. This difference is presumably because none of these systems has a need for more than one ACP to engage the KS active site. The dual KS entrances may also facilitate module 'skipping', as observed in the Pik and DEBS systems[43,44].

The PikAIII dimeric architecture is fundamental to bacterial type I modular PKSs because it creates a single ACP chamber. Thus, it seems odd that the structures of excised KS–AT di-domains[6,7,21] fail to mimic the structure of the full module. Perhaps the answer lies in the fact that all PKS modules have at least one dimerization element that is C-terminal to the AT domain and is, therefore, missing from excised KS–AT di-domains[4,45]. Dimerization can occur through a dehydratase domain[12,13], post-ACP dimerization helices, as in PikAIII[4], or a post-AT dimerization domain[21,45]. With respect to bimodules that are naturally fused, the downstream KS is a dimerization element for the upstream module because the intermodule linker has neither post-ACP dimerization helices nor a pre-KS coiled coil. Furthermore, in PikAIII, the post-ACP dimerization helices are essential to dimer formation (Extended Data Fig. 9a) and exert an outsize influence on the overall module structure, as seen in the large rotation of the KR_5 (165°) upon ACP_5 deletion (PikAIII (ΔACP_5)) (Extended Data Fig. 7d) and the in-concert movement of the two ACP_5 domains.

The ACP_5 position within the module appears to be dictated by the substrate linked to the phosphopantetheine arm, as indicated by the specific ACP localization in the MM–PikAIII state and in the ACP_4–PikAIII fusion loaded with pentaketide. The surfaces of ACP_4 and ACP_5 that engage the enzyme domains are similar to those involved in formation of other ACP–enzyme complexes[46–48]. These findings provide the first evidence of an elegant mechanistic solution for ACP transfer whereby the identity of the processed substrate itself, instead of specific protein–protein interactions, is the major determinant of ACP positioning, consistent with observations that natural intermediates are processed more effectively than unnatural substrates[24,49]. This brilliant design allows for efficient and specific transfer of intermediates and optimal throughput of the assembly line.

METHODS SUMMARY

PikAIII variants for cryo-EM analysis were purified by Ni-affinity chromatography followed by duplicate gel filtration steps. To trap PikAIII in different states, the purified proteins were incubated with substrate, and adsorbed on holey carbon grids followed by vitrification. Bottom-up LC/FT-ICR MS confirmed substrate loading in each of the PikAIII states. All specimens were imaged on a Tecnai F20 transmission electron microscope (FEI) equipped with a field emission gun operated at 120 kV. Images were recorded at a magnification of ×66,964 on a Gatan US4000 CCD camera with defocus values ranging from −1.5 to −3.5 μm. EMAN routines were used for interactive particle selection, generation of initial models, and three-dimensional map refinement. Crystal structures of domains were independently fit in the EM maps as rigid bodies. Functional assays that require intermediate processing by PikAIII detected TE-offloaded products from PikAIII–TE or PikAIII/PikAIV[24,25].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 May 2013; accepted 1 May 2014.
Published online 18 June 2014.

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75,** 311–335 (2012).
2. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106,** 3468–3496 (2006).
3. Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24,** 1041–1072 (2007).
4. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10,** 723–731 (2003).

©2014 Macmillan Publishers Limited. All rights reserved

5. Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4,** 41–52 (2009).
6. Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14,** 931–943 (2007).
7. Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl Acad. Sci. USA* **103,** 11124–11129 (2006).
8. Zheng, J. & Keatinge-Clay, A. T. Structural and functional analysis of C2-type ketoreductases from modular polyketide synthases. *J. Mol. Biol.* **410,** 105–117 (2011).
9. Zheng, J., Taylor, C. A., Piasecki, S. K. & Keatinge-Clay, A. T. Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* **18,** 913–922 (2010).
10. Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14,** 898–908 (2007).
11. Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. *Structure* **14,** 737–748 (2006).
12. Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384,** 941–953 (2008).
13. Akey, D. L. *et al.* Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18,** 94–105 (2010).
14. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nature Chem. Biol.* **8,** 615–621 (2012).
15. Gehret, J. J. *et al.* Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. *J. Biol. Chem.* **286,** 14445–14454 (2011).
16. Scaglione, J. B. *et al.* Biochemical and structural characterization of the tautomycetin thioesterase: analysis of a stereoselective polyketide hydrolase. *Angew. Chem. Int. Edn Engl.* **49,** 5726–5730 (2010).
17. Tsai, S. C., Lu, H., Cane, D. E., Khosla, C. & Stroud, R. M. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **41,** 12598–12606 (2002).
18. Tsai, S. C. *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc. Natl Acad. Sci. USA* **98,** 14808–14813 (2001).
19. Bonnett, S. A. *et al.* Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a *cis*-alkene. *Chem. Biol.* **20,** 772–783 (2013).
20. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain–domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16,** 2093–2107 (2007).
21. Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains, a new tool for engineering natural product biosynthesis. *Chem. Biol.* **20,** 1340–1351 (2013).
22. Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321,** 1315–1322 (2008).
23. Xue, Y., Zhao, L., Liu, H. W. & Sherman, D. H. A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proc. Natl Acad. Sci. USA* **95,** 12111–12116 (1998).
24. Aldrich, C. C., Beck, B. J., Fecik, R. A. & Sherman, D. H. Biochemical investigation of pikromycin biosynthesis employing native penta- and hexaketide chain elongation intermediates. *J. Am. Chem. Soc.* **127,** 8441–8452 (2005).
25. Aldrich, C. C., Venkatraman, L., Sherman, D. H. & Fecik, R. A. Chemoenzymatic synthesis of the polyketide macrolactone 10-deoxymethynolide. *J. Am. Chem. Soc.* **127,** 8910–8911 (2005).
26. Lyon, A. M., Dutta, S., Boguth, C. A., Skiniotis, G. & Tesmer, J. J. Full-length $G\alpha_q$–phospholipase C-β3 structure reveals interfaces of the C-terminal coiled-coil domain. *Nature Struct. Mol. Biol.* **20,** 355–362 (2013).
27. Strunk, B. S. *et al.* Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science* **333,** 1449–1453 (2011).
28. Staunton, J. *et al.* Evidence for a double-helical structure for modular polyketide synthases. *Nature Struct. Biol.* **3,** 188–192 (1996).
29. Kao, C. M., Pieper, R., Cane, D. E. & Khosla, C. Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. *Biochemistry* **35,** 12363–12368 (1996).
30. Gokhale, R. S., Lau, J., Cane, D. E. & Khosla, C. Functional orientation of the acyltransferase domain in a module of the erythromycin polyketide synthase. *Biochemistry* **37,** 2524–2528 (1998).
31. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* http://dx.doi.org/10.1038/nature13409 (this issue).
32. Yin, Y., Lu, H., Khosla, C. & Cane, D. E. Expression and kinetic analysis of the substrate specificity of modules 5 and 6 of the picromycin/methymycin polyketide synthase. *J. Am. Chem. Soc.* **125,** 5671–5676 (2003).
33. Yuzawa, S., Kapur, S., Cane, D. E. & Khosla, C. Role of a conserved arginine residue in linkers between the ketosynthase and acyltransferase domains of multimodular polyketide synthases. *Biochemistry* **51,** 3708–3710 (2012).
34. Kittendorf, J. D., Beck, B. J., Buchholz, T. J., Seufert, W. & Sherman, D. H. Interrogating the molecular basis for multiple macrolactone ring formation by the pikromycin polyketide synthase. *Chem. Biol.* **14,** 944–954 (2007).
35. Kapur, S., Chen, A. Y., Cane, D. E. & Khosla, C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl Acad. Sci. USA* **107,** 22066–22071 (2010).
36. Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl Acad. Sci. USA* **109,** 4110–4115 (2012).
37. Bonnett, S. A. *et al.* Acyl-CoA subunit selectivity in the pikromycin polyketide synthase PikAIV: steady-state kinetics and active-site occupancy analysis by FTICR-MS. *Chem. Biol.* **18,** 1075–1081 (2011).
38. Huang, W. *et al.* Crystal structure of β-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. *EMBO J.* **17,** 1183–1191 (1998).
39. Chemler, J. A. *et al.* Biochemical and structural characterization of germicidin synthase: analysis of a type III polyketide synthase that employs acyl-ACP as a starter unit donor. *J. Am. Chem. Soc.* **134,** 7359–7366 (2012).
40. Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C. & Stroud, R. M. An antibiotic factory caught in action. *Nature Struct. Biol.* **11,** 888–893 (2004).
41. Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A. & Noel, J. P. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Struct. Biol.* **6,** 775–784 (1999).
42. Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316,** 288–290 (2007).
43. Beck, B. J., Yoon, Y. J., Reynolds, K. A. & Sherman, D. H. The hidden steps of domain skipping: macrolactone ring size determination in the pikromycin modular polyketide synthase. *Chem. Biol.* **9,** 575–583 (2002).
44. Rowe, C. J. *et al.* Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem. Biol.* **8,** 475–485 (2001).
45. Zheng, J., Fage, C. D., Demeler, B., Hoffman, D. W. & Keatinge-Clay, A. T. The missing linker: a dimerization motif located within polyketide synthase modules. *ACS Chem. Biol.* **8,** 1263–1270 (2013).
46. Bunkoczi, G. *et al.* Mechanism and substrate recognition of human holo ACP synthase. *Chem. Biol.* **14,** 1243–1253 (2007).
47. Masoudi, A., Raetz, C. R., Zhou, P. & Pemble, C. W. IV Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature* **505,** 422–426 (2014).
48. Nguyen, C. *et al.* Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505,** 427–431 (2014).
49. Wu, N., Tsuji, S. Y., Cane, D. E. & Khosla, C. Assessing the balance between protein–protein interactions and enzyme–substrate interactions in the channeling of intermediates between polyketide synthase modules. *J. Am. Chem. Soc.* **123,** 6465–6474 (2001).

Acknowledgements We thank R. Henderson for sharing the phase-randomization program and for valuable advice on the implementation of the tilt-pair validation, and Y. Cheng and X. Li for sharing the image sub-frame alignment program. This work was supported by the Pew Scholar Program in Biomedical Sciences (G.S.), the University of Michigan Biological Sciences Scholars Program (G.S.), Rackham Merit and American Foundation for Pharmaceutical Education predoctoral fellowships (D.A.H.), a National Research Service Award postdoctoral fellowship (J.A.C.), the Life Sciences Research Foundation (A.R.H.N.), National Institutes of Health grants 1R21CA138331-01A1 (K.H.), GM076477 (D.H.S. and J.L.S.), DK042303 (J.L.S.) and DK090165 (G.S.), and the Hans W. Vahlteich Professorship (to D.H.S.).

Author Contributions S.D. carried out cryo-EM data collection and processing. J.R.W. produced PikAIII variants and conducted enzyme assays. G.R.C. assisted with cryo-EM image processing. W.A.H., A.R.H.N. and K.H. carried out mass-spectrometry analysis. D.A.H. synthesized the thiophenol-activated pentaketide and pentaketide–CoA substrates. J.A.C. produced initial PikAIII samples and prepared Fig. 1. S.D., J.R.W., J.L.S. and G.S. analysed the data and interpreted results. K.H., D.H.S., J.L.S. and G.S. designed research. S.D., J.R.W., D.H.S., J.L.S. and G.S. wrote the manuscript.

Author Information Cryo-EM maps have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-5647 (holo-PikAIII conformation I), EMD-5648 (holo-PikAIII conformation II), EMD-5649 (PikAIII(ΔACP_5)), EMD-5651 (pentaketide–ACP_4–PikAIII(C209A/ΔACP_5)), EMD-5653 (MM–PikAIII) and EMD-5662 (holo-ACP_4–PikAIII(C209A/ΔACP_5)). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.S. (skinioti@umich.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13433

Precision measurement of the Newtonian gravitational constant using cold atoms

G. Rosi[1], F. Sorrentino[1], L. Cacciapuoti[2], M. Prevedelli[3] & G. M. Tino[1]

About 300 experiments have tried to determine the value of the Newtonian gravitational constant, *G*, so far, but large discrepancies in the results have made it impossible to know its value precisely[1]. The weakness of the gravitational interaction and the impossibility of shielding the effects of gravity make it very difficult to measure *G* while keeping systematic effects under control. Most previous experiments performed were based on the torsion pendulum or torsion balance scheme as in the experiment by Cavendish[2] in 1798, and in all cases macroscopic masses were used. Here we report the precise determination of *G* using laser-cooled atoms and quantum interferometry. We obtain the value $G = 6.67191(99) \times 10^{-11}\ \mathrm{m^3\,kg^{-1}\,s^{-2}}$ with a relative uncertainty of 150 parts per million (the combined standard uncertainty is given in parentheses). Our value differs by 1.5 combined standard deviations from the current recommended value of the Committee on Data for Science and Technology[3]. A conceptually different experiment such as ours helps to identify the systematic errors that have proved elusive in previous experiments, thus improving the confidence in the value of *G*. There is no definitive relationship between *G* and the other fundamental constants, and there is no theoretical prediction for its value, against which to test experimental results. Improving the precision with which we know *G* has not only a pure metrological interest, but is also important because of the key role that *G* has in theories of gravitation, cosmology, particle physics and astrophysics and in geophysical models.

The basic idea of our experiment is to use an atom interferometer as a gravity sensor and a well-characterized mass as the source of a gravitational field. From the precise measurement of the atoms' acceleration produced by the source mass and from the knowledge of the mass distribution, it is possible to extract the value of *G* using the formula

$$F(r) = -G\frac{M_1 M_2}{r^2}\hat{r}$$

where $\hat{r}$ is the radial unit vector.

Atom interferometers[4,5] are new tools for experimental gravitation, for example in precision measurements of gravitational acceleration[6] and gravity gradients[7], as gyroscopes based on the Sagnac effect[8], for testing the $1/r^2$ law[9], in general relativity[10] and quantum gravity models[11], and in applications in geophysics[12]. Proof-of-principle experiments to measure *G* using atom interferometry have been reported[13–15]. Ongoing studies show that future experiments in space will take full advantage of the potential sensitivity of atom interferometers for fundamental physics tests[16]. The possibility of using atom interferometry for gravitational wave detection is being studied[17].

Because the problem in the determination of *G* depends on the presence of unidentified systematic errors, our experiment was designed with a double-differential configuration to be as insensitive as possible to such effects: the atomic sensor was a double interferometer in a gravity gradiometer configuration, to subtract common-mode spurious signals, and to produce the gravitational field we used two sets of well-characterized tungsten masses that were placed in two different positions to modulate the relevant gravitational signal. An additional cancellation of common-mode spurious effects was obtained by reversing the direction of the two-photon recoil used to split and recombine the wave packets in the interferometer[18]. Efforts were devoted to the control of systematics related to atomic trajectories, the positioning of the atoms and effects due to stray fields. The high density of tungsten was instrumental in maximizing the signal and in compensating for the Earth's gravitational gradient in the region containing the atom interferometers, thus reducing the sensitivity of the experiment to the vertical position and size of the atomic probes.

The atom interferometer is realized using light pulses to stimulate ^{87}Rb atoms at the two-photon Raman transition between the hyperfine levels $F = 1$ and $F = 2$ of the ground state[19]. The light field is generated by two counter-propagating laser beams with wave vectors k_1 and $k_2 \approx -k_1$ aligned along the vertical direction. The gravity gradiometer consists of two vertically separated atom interferometers operated in differential mode. Two atomic clouds launched along the vertical direction are simultaneously interrogated by the same $\pi/2$–π–$\pi/2$ pulse sequence. The difference in the phase shifts detected at the output of each interferometer provides a direct measurement of the differential acceleration induced by gravity on the two atomic samples. In this way, any spurious acceleration induced by vibrations or seismic noise in the common reference frame identified by the vertical Raman beams is efficiently rejected.

Figure 1 shows a sketch of the experiment. The atom interferometer apparatus and the source mass assembly are described in detail elsewhere[20,21]. In the vacuum chamber at the bottom of the apparatus, a magneto-optical trap (MOT) collects $\sim 10^9$ rubidium atoms. After turning the MOT magnetic field off, the atoms are launched vertically along the symmetry axis of the vacuum tube by using the 'moving-molasses' technique. During the launch sequence, atoms are laser cooled to a temperature of $\sim 4\ \mu$K. We juggle two atomic samples to have them reach the apogees of their ballistic trajectories at about 60 and, respectively, 90 cm above the MOT, with a vertical separation of 328 mm.

The atoms are velocity-selected and prepared in the magnetic-field-insensitive $|F = 1, m_F = 0\rangle$ hyperfine state with a combination of three π Raman pulses and two resonant laser pulses used to remove the atoms occupying the wrong hyperfine state. The interferometers are realized at the centre of the vertical tube shown in Fig. 1. In this region, surrounded by two cylindrical magnetic shields, a uniform magnetic field of 29 μT along the vertical direction defines the quantization axis. Here atoms are subjected to the Raman three-pulse interferometer sequence. The central π pulse occurs about 6 ms after the atoms reach the apogees of their trajectories. At the end of the atoms' ballistic flight, the population of the ground state is measured by selectively exciting the atoms in both hyperfine levels of the ground state and detecting the light-induced fluorescence emission. We typically detect 10^5 atoms on each rubidium sample at the end of the interferometer sequence. Each measurement takes 1.9 s. The information on the relative phase shift between the two atom interferometers is extracted from the Lissajous curve that is obtained when the signal of one interferometer is plotted as a function of the signal

[1]Dipartimento di Fisica e Astronomia and LENS, Università di Firenze - INFN Sezione di Firenze, Via Sansone 1, 50019 Sesto Fiorentino, Italy. [2]European Space Agency, Keplerlaan 1, PO Box 299, 2200 AG Noordwijk ZH, The Netherlands. [3]Dipartimento di Fisica, Università di Bologna, Via Irnerio 46, 40126, Bologna, Italy.

©2014 Macmillan Publishers Limited. All rights reserved

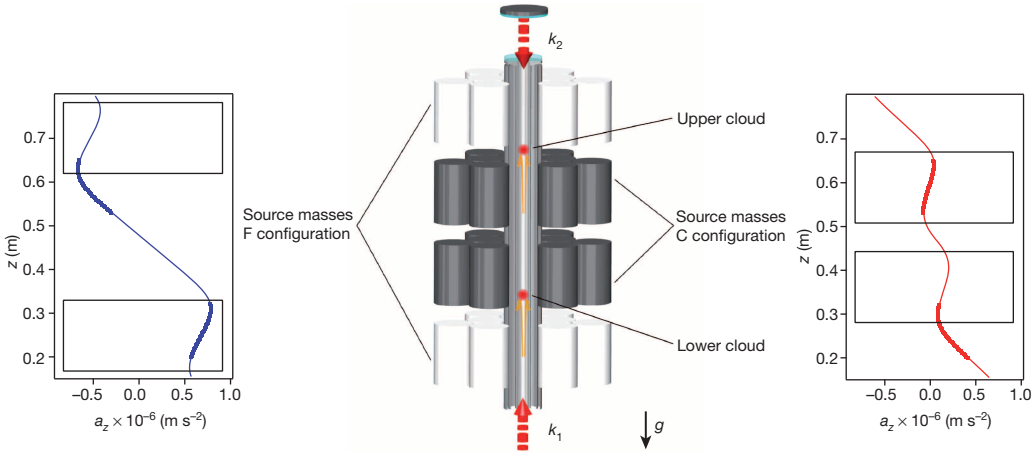


Figure 1 | Sketch of the experiment. The Rb atom interferometer operates as a gravity gradiometer and the W masses are used as the source of the gravitational field. For the measurement of G, the position of the source masses is alternated between configurations F and C. Plots of gravitational acceleration (a_z) produced along the symmetry axis by the source masses are also shown for each configuration; a constant value for Earth's gravity was subtracted. The spatial regions of the upper and lower atom interferometers are indicated by the thick lines. The vertical acceleration plots show the effect of source mass in cancelling the local gravity gradient at the positions of the atomic apogees.

of the other. Experimental points are distributed around an ellipse. The differential phase shift is extracted from the eccentricity and the rotation angle of the ellipse fitting the data[22]. The instrument sensitivity for differential acceleration is $3 \times 10^{-9} g$ for 1 s of integration (g is the acceleration due to Earth's gravity).

The source mass is composed of 24 tungsten alloy cylinders, for a total mass of about 516 kg. Each cylinder is machined to have a diameter of 99.90 mm and a height of 150.11 mm. They are positioned on two titanium platforms and distributed with hexagonal symmetry around the vertical axis of the tube (Fig. 1). The cylinders' centres lie around two circles with nominal radii $2R$ and $2R\sqrt{3}$, respectively, where R is the radius of a single cylinder. The vertical positioning of the two platforms is ensured by precision screws synchronously driven by stepper motors and by an optical readout system. The reproducibility of the positioning was verified with a laser tracker to be within 1 μm (ref. 20). With respect to the position of the apogee of the lower atomic cloud, the centres of the lower and upper sets of cylinders lie at respective vertical distances of 40 and 261 mm in one configuration (the C configuration) and at −74 and 377 mm in another (the F configuration).

The value of the Newtonian gravitational constant was obtained from a series of gravity gradient measurements performed by periodically changing the vertical position of the source masses between configurations

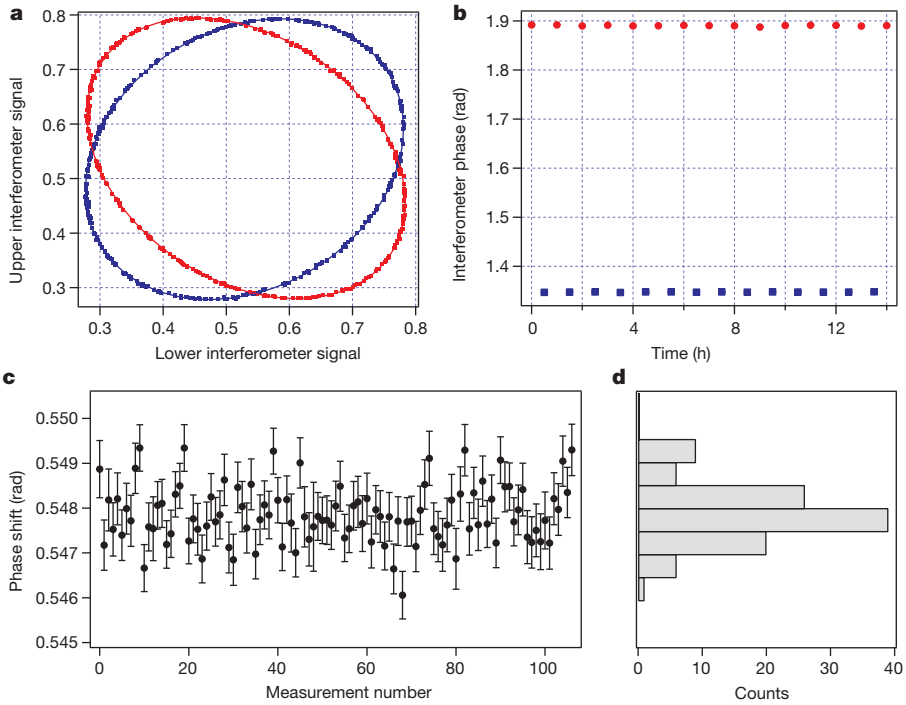


Figure 2 | Experimental data. **a**, Typical Lissajous figures obtained by plotting the output signal of the upper atom interferometer versus that of the lower one for the two configurations of the source masses. **b**, Modulation of the differential phase shift for the two configurations of source masses for a given direction of the Raman beams' k vector. Each phase measurement is obtained by fitting a 360-point scan of the atom interference fringes to an ellipse. The error bars, not visible on this scale, are given by the standard error of the least-squares fit to the ellipse. **c**, Results of the measurements to determine G. Each point is obtained by averaging the signals recorded for the two directions of the Raman k vector (Methods). Data acquisition for each point took about one hour. These data were recorded on different days during one week in July 2013. The error bars are given by the combined errors in the phase angles of four ellipses. **d**, Histogram of the data in **c**.

©2014 Macmillan Publishers Limited. All rights reserved

Table 1 | Effects, relative corrections and uncertainties considered in our determination of *G*

Parameter	Uncertainty in parameter	Relative correction to *G* (p.p.m.)	Relative uncertainty in *G* (p.p.m.)
Air density	10%	60	6
Apogee time	30 μs	—	6
Atomic cloud horizontal size	0.5 mm	—	24
Atomic cloud vertical size	0.1 mm	—	56
Atomic cloud horizontal position	1 mm	—	37
Atomic cloud vertical position	0.1 mm	—	5
Atom launch direction change C/F	8 μrad	—	36
Cylinder density homogeneity	10^{-4}	91	18
Cylinder radial position	10 μm	—	38
Ellipse fit	—	−13	4
Size of detection region	1 mm	—	13
Support platform mass	10 g	—	5
Translation stage position	0.5 mm	—	6
Other effects	—	<2	1
Total systematic uncertainty	—	—	92
Statistical uncertainty	—	—	116
Total	—	137	148

Uncertainties are quoted as one standard deviation. The third column contains the corrections we applied to account for effects not included in the Monte Carlo simulation. The bias and systematic error from ellipse fitting are evaluated by a numerical simulation on synthetic data. Other effects include cylinder mass, cylinder vertical position, gravity gradient, gravity acceleration, Raman mirror tilt, Raman *k* vector and timing.

C and F. Figure 2 shows the data used for the determination of *G*. Data were collected in 100 h during one week in July 2013. Each phase measurement was obtained by fitting a 360-point scan of the atom interference fringes to an ellipse. The modulation of the differential phase shift produced by the source mass is easily visible and could be resolved with a signal-to-noise ratio of 1,000 after about one hour. The resulting value of the differential phase shift is 0.547870(63) rad, and it was from this that we obtained *G*. The cylinders produce 97.0% of the measured differential phase shift, the cylinder supports produce 2.8% (ref. 20) and the additional moving masses (translation stages, optical rulers, screws) produce the remaining 0.2%.

The sources of uncertainty affecting the value of *G* are presented in Table 1. Positioning errors account for uncertainties in the positions of the 24 tungsten cylinders along the radial and vertical direction, both in configuration C and configuration F. Density inhomogeneities in the source masses were measured by cutting and weighing a spare cylinder[20], and were modelled in the data analysis. Precise knowledge of the atomic trajectories is of key importance in analysing the experimental results and deriving the value of *G*. The velocities of the atomic clouds, and their positions at the time of the first interferometer pulse, were calibrated by time-of-flight measurements and by detecting the atoms when they crossed a horizontal light sheet while moving upwards and downwards. The Earth's rotation affects the atom interferometers' signals because of the transverse velocity distribution of the atoms. Following the method demonstrated for a single interferometer[23,24], we implemented a tip–tilt scheme for the mirror retroreflecting the Raman beams in our double interferometer.

Extracting the value of *G* from the data involved the following steps: calculation of the gravitational potential produced by the source masses; calculation of the phase shift for single-atom trajectories; Monte Carlo simulation of the atomic cloud; and calculation of the corrections for the effects not included in the Monte Carlo simulation (Table 1).

After an analysis of the error sources affecting our measurement, we obtain the value $G = 6.67191(77)(62) \times 10^{-11}\ \mathrm{m^3\,kg^{-1}\,s^{-2}}$. The statistical and systematic errors, reported in parenthesis as one standard deviation, lead to a combined relative uncertainty of 150 p.p.m. In Fig. 3, this result is compared with the values of recent experiments and Committee on Data for Science and Technology (CODATA) adjustments. Our value,

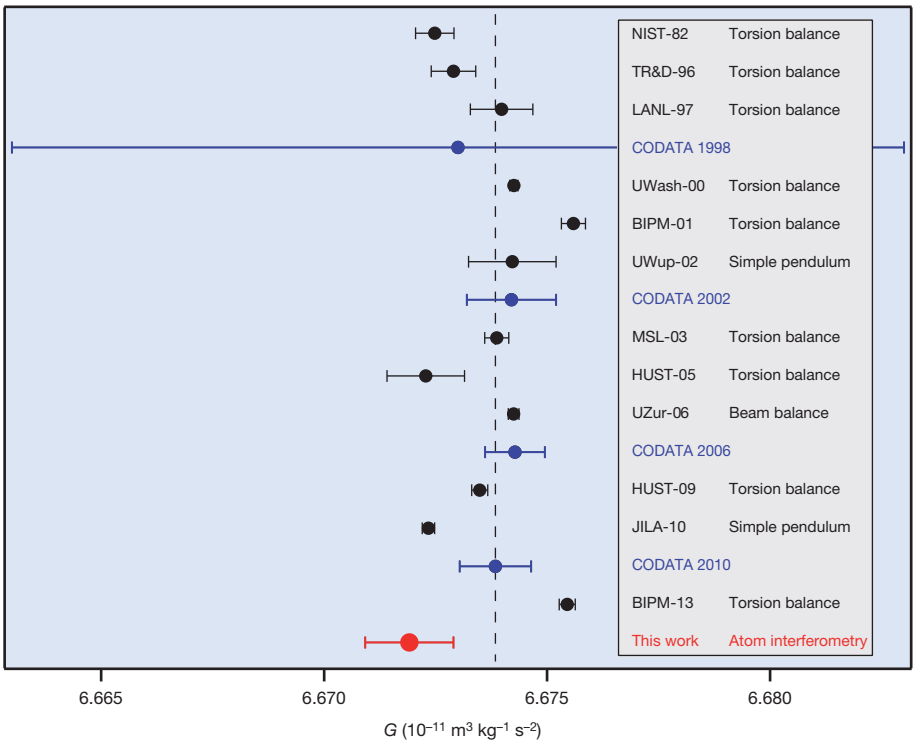


Figure 3 | Comparison with previous results. Result of this experiment for *G* compared with the values obtained in previous experiments and with the recent CODATA adjustments. Only the experiments considered for the current CODATA 2010 value, and the subsequent BIPM-13 result, are included. For details on the experiments and their identification with the abbreviations used in the figure, see ref. 3 and the additional references in Methods.

©2014 Macmillan Publishers Limited. All rights reserved

obtained with a method completely different from the ones used before, differs by 1.5 combined standard deviations from the present CODATA value, $G = 6.67384(80) \times 10^{-11}\ \mathrm{m^3\,kg^{-1}\,s^{-2}}$.

Conceptually different experiments such as ours will help to identify the errors that produced the remaining discrepancies among previous measurements. Given the relevance of the gravitational constant in several fields ranging from cosmology to particle physics, and in the absence of a complete theory linking gravity to other forces, high-precision measurements based on different methods are crucial to increase our confidence in the value of G. It is possible to imagine a new generation of experiments to measure G using ultracold atoms confined in optical traps to allow their precise positioning. The remaining major contribution to the systematic uncertainty in our experiment in fact derives from the positioning of the atoms with respect to the source mass. The choice of a different atom might reduce this source of uncertainty as well as others. For example, Sr atoms can be rapidly cooled to form Bose–Einstein condensates[25], they are virtually unaffected by magnetic fields and collisional effects in their ground state, and the possibility of their precise positioning and use in gravity measurements using optical lattices has been demonstrated[26,27]. Higher-sensitivity atom interferometers with larger splitting of the atomic wave packet would enable the use of a smaller, highly homogeneous source mass such as gold or, eventually, silicon crystals. We foresee prospects for increasing the measurement accuracy in the determination of G to better than 10 p.p.m. using atoms.

METHODS SUMMARY

The Raman lasers frequencies ν_1 and ν_2 satisfy the resonance condition $\nu_2 - \nu_1 \approx 6.8$ GHz, corresponding to the separation between the $F = 1$ and $F = 2$ levels of ^{87}Rb ground state. During absorption and stimulated emission processes, the resonant light field exchanges with atoms a total momentum of $\hbar k = \hbar k_1 - \hbar k_2$ ($\hbar$ is Planck's constant divided by 2π), coupling the two states $|F = 1, p\rangle$ and $|F = 2, p + \hbar k\rangle$, where p is the initial atomic momentum. The interferometer is composed of a $\pi/2$–π–$\pi/2$ sequence of Raman pulses separated by a time $T = 160$ ms. The pulses split, redirect and recombine the wave packets, producing atom interference. At the end of the sequence, the probability of detecting atoms in $|F = 2, p + \hbar k\rangle$ is $P_2 = [1 - \cos\phi]/2$, where ϕ is the phase difference accumulated along the interferometer arms. In a uniform gravitational field, atoms experience a phase shift $\phi = k \bullet gT^2$ that can be measured and thus provides direct information on the local acceleration due to gravity. Accurate control of noise sources and systematic phase shifts is crucial to optimize the sensitivity and the long-term stability of gravity gradient measurements[28]. Active loops have been implemented to stabilize the optical intensities of the cooling, Raman and probe laser beams and the Raman mirror tilt. In this way, we reach a quantum-projection-noise-limited sensitivity to differential accelerations of $3 \times 10^{-9} g$ after 1 s of integration time and an accuracy in the measurement of G at the level of 100 p.p.m. The measured phase shift is compared with the theoretical value obtained by evaluating the action integral along the trajectories given by the solution of the Lagrange equation:

$$\phi = \frac{1}{\hbar}\oint L\,\mathrm{d}t \qquad (1)$$

Taking a perturbative approach[29], we separate the Lagrangian L into $L_0 = p^2/2m - mgz$ (m is the atomic mass) and L_1, which accounts for the contributions of both the Earth's gravity gradient and the source mass. A Monte Carlo simulation is used to evaluate the measurement results and to determine the sensitivity of the final phase angle to the relevant parameters. To this end, the key parameters entering the simulation are varied and their derivatives calculated to estimate the uncertainty in the measurement of G.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 January; accepted 22 April 2014.
Published online 18 June 2014.

1. Quinn, T. Measuring big *G*. *Nature* **408,** 919–921 (2000).
2. Cavendish, H. Experiments to determine the density of the Earth. *Phil. Trans. R. Soc. Lond.* **88,** 469–526 (1798).
3. Mohr, P. J., Taylor, B. N. & Newell, D. B. CODATA recommended values of the fundamental physical constants: 2010. *Rev. Mod. Phys.* **84,** 1527–1605 (2012).
4. Cronin, A. D., Schmiedmayer, J. & Pritchard, D. E. Optics and interferometry with atoms and molecules. *Rev. Mod. Phys.* **81,** 1051–1129 (2009).
5. Tino, G. M. & Kasevich, M. A. (eds) *Atom Interferometry, Proc. Int. School Phys. "Enrico Fermi", Course CLXXXVIII, Varenna 2013* (Società Italiana di Fisica and IOS Press, 2014).
6. Peters, A., Chung, K. Y. & Chu, S. Measurement of gravitational acceleration by dropping atoms. *Nature* **400,** 849–852 (1999).
7. McGuirk, J. M., Foster, G. T., Fixler, J. B., Snadden, M. J. & Kasevich, M. A. Sensitive absolute-gravity gradiometry using atom interferometry. *Phys. Rev. A* **65,** 033608 (2002).
8. Gustavson, T. L., Landragin, A. & Kasevich, M. A. Rotation sensing with a dual atom interferometer Sagnac gyroscope. *Class. Quantum Gravity* **17,** 2385–2398 (2000).
9. Ferrari, G., Poli, N., Sorrentino, F. & Tino, G. M. Long-lived Bloch oscillations with bosonic Sr atoms and application to gravity measurement at the micrometer scale. *Phys. Rev. Lett.* **97,** 060402 (2006).
10. Dimopoulos, S., Graham, P., Hogan, J. & Kasevich, M. Testing general relativity with atom interferometry. *Phys. Rev. Lett.* **98,** 111102 (2007).
11. Amelino-Camelia, G., Lämmerzahl, C., Mercati, F. & Tino, G. M. Constraining the energy-momentum dispersion relation with Planck-scale sensitivity using cold atoms. *Phys. Rev. Lett.* **103,** 171302 (2009).
12. de Angelis, M. *et al.* Precision gravimetry with atomic sensors. *Meas. Sci. Technol.* **20,** 022001 (2009).
13. Bertoldi, A. *et al.* Atom interferometry gravity-gradiometer for the determination of the Newtonian gravitational constant *G*. *Eur. Phys. J. D* **40,** 271–279 (2006).
14. Fixler, J. B., Foster, G. T., McGuirk, J. M. & Kasevich, M. Atom interferometer measurement of the Newtonian constant of gravity. *Science* **315,** 74–77 (2007).
15. Lamporesi, G., Bertoldi, A., Cacciapuoti, L., Prevedelli, M. & Tino, G. M. Determination of the Newtonian gravitational constant using atom interferometry. *Phys. Rev. Lett.* **100,** 050801 (2008).
16. Tino, G. M. *et al.* Precision gravity tests with atom interferometry in space. *Nucl. Phys. B Proc.* Suppl. **243–244,** 203–217 (2013).
17. Tino, G. M., Vetrano, F. & Lämmerzahl, C. Editorial on the special issue on "gravitational waves detection with atom interferometry". *Gen. Relativ. Gravit.* **43,** 1901–1903 (2011).
18. Louchet-Chauvet, A. *et al.* The influence of transverse motion within an atomic gravimeter. *New J. Phys.* **13,** 065025 (2011).
19. Kasevich, M. & Chu, S. Atomic interferometry using stimulated Raman transitions. *Phys. Rev. Lett.* **67,** 181–184 (1991).
20. Lamporesi, G. *et al.* Source masses and positioning system for an accurate measurement of *G*. *Rev. Sci. Instrum.* **78,** 075109 (2007).
21. Sorrentino, F. *et al.* Sensitive gravity-gradiometry with atom interferometry: progress towards an improved determination of the gravitational constant. *New J. Phys.* **12,** 095009 (2010).
22. Foster, G. T., Fixler, J. B., McGuirk, J. M. & Kasevich, M. A. novel method of phase extraction between coupled atom interferometers using ellipse-specific fitting. *Opt. Lett.* **27,** 951–953 (2002).
23. Hogan, J., Johnson, D. & Kasevich, M. in *Proc. Int. School Phys. "Enrico Fermi", Course CLXVIII* (eds Arimondo, E., Ertmer, W., Schleich, W. & Rasel, E.) 411–447 (SIF, Bologna and IOS Press, 2009).
24. Lan, S.-Y., Kuan, P.-C., Estey, B., Haslinger, P. & Müller, H. Influence of the Coriolis force in atom interferometry. *Phys. Rev. Lett.* **108,** 090402 (2012).
25. Stellmer, S., Pasquiou, B., Grimm, R. & Schreck, F. Laser cooling to quantum degeneracy. *Phys. Rev. Lett.* **110,** 263003 (2013).
26. Sorrentino, F. *et al.* Quantum sensor for atom-surface interactions below 10 μm. *Phys. Rev. A* **79,** 013409 (2009).
27. Poli, N. *et al.* Precision measurement of gravity with cold atoms in an optical lattice and comparison with a classical gravimeter. *Phys. Rev. Lett.* **106,** 038501 (2011).
28. Sorrentino, F. *et al.* Sensitivity limits of a Raman atom interferometer as a gravity gradiometer. *Phys. Rev. A* **89,** 023607 (2014).
29. Storey, P. & Cohen-Tannoudji, C. The Feynman path integral approach to atomic interferometry. A tutorial. *J. Phys. II France* **4,** 1999–2027 (1994).

Acknowledgements G.M.T. acknowledges discussions with M. A. Kasevich and J. Faller and useful suggestions by A. Peters in the initial phase of the experiment. We are grateful to A. Cecchetti and B. Dulach for the design of the source mass support and to A. Peuto, A. Malengo, and S. Pettorruso for density tests on the tungsten masses. We thank D. Wiersma for a critical reading of the manuscript. This work was supported by INFN (MAGIA experiment).

Author Contributions G.M.T. had the idea for the experiment, supervised it and wrote the manuscript. G.R., F.S. and L.C. performed the experiment. M.P. contributed to the experiment and analysed the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.M.T. (guglielmo.tino@fi.infn.it).

©2014 Macmillan Publishers Limited. All rights reserved

METHODS

Experimental set-up. Our interferometer uses Raman pulses to drive ^{87}Rb atoms on the two-photon transition between the hyperfine levels $|1\rangle \equiv |F=1, m_F=0\rangle$ and $|2\rangle \equiv |F=2, m_F=0\rangle$ of the ground state[19]. The light field is generated by two counter-propagating laser beams with wave vectors k_1 and $k_2 \approx -k_1$ aligned along the vertical direction. This configuration is obtained by retroreflecting the Raman lasers using a mirror placed on top of the vertical tube (Fig. 1). The laser frequencies ν_1 and ν_2 satisfy the resonance condition $\nu_2 - \nu_1 = \nu_0 + \nu_{AC} + \nu_{rec}$, where $h\nu_0$ is the energy corresponding to the $|1\rangle \rightarrow |2\rangle$ transition, $h\nu_{AC}$ accounts for the a.c. Stark effect and $h\nu_{rec}$ is the energy shift due to the two-photon recoil. As a consequence, the transition modifies both the internal energy and the total momentum of the atom, coupling the two atomic states $|1, p\rangle$ and $|2, p + \hbar k\rangle$, where $k = k_1 - k_2$ is the effective wavevector of the Raman transition. The $\hbar k$ momentum transfer is responsible for the spatial separation of the atomic wave packet along the two physical arms of the interferometer. During interferometry, a sequence of three Raman pulses provides the atom-optical elements of a typical Mach–Zehnder configuration: at $t = 0$ a $\pi/2$ beam-splitter pulse prepares the atoms, initially in $|1, p\rangle$, in an equal and coherent superposition of $|1, p\rangle$ and $|2, p + \hbar k\rangle$; after a time $T = 160$ ms a π pulse plays the part of a mirror, swapping state $|1, p\rangle$ with $|2, p + \hbar k\rangle$ and redirecting the atoms towards the output ports of the interferometer; at $t = 2T$ a final $\pi/2$ pulse recombines the atomic wave packets in the two output ports, where they interfere. When interacting with the light field, the phase of the Raman lasers is imprinted on the atomic wavefunction, sampling the atomic motion at the three interaction events of the $\pi/2$–π–$\pi/2$ sequence. This information can be read through the effect of matter-wave interference by detecting the atomic population in the $F = 1$ and $F = 2$ hyperfine states. If ϕ is the phase difference accumulated by the atoms along the two interferometer arms, the probability of detecting atoms in $|2\rangle$ can be expressed as $P_2 = [1 - \cos\phi]/2$. In the presence of a gravity field, atoms experience a phase shift $\phi = k \bullet gT^2$. Measuring ϕ is therefore equivalent to a measurement of the local acceleration due to the gravitational field along the direction of the effective wavevector, k.

Our instrument is a gravity gradiometer consisting of two Raman-pulse interferometers simultaneously probing two clouds of laser-cooled ^{87}Rb atoms aligned along the vertical axis and separated by a distance of 328 mm (refs 21, 28). The two atomic clouds are interrogated by the same Raman lasers with the same interferometric pulse sequence. The difference in the phase shifts at each interferometer provides a direct measurement of the differential acceleration induced by gravity on the two atomic samples. This configuration allows a very efficient rejection of common-mode noise sources, including spurious accelerations at the instrument platform induced by mechanical vibrations or seismic noise, g variations due to tidal effects and phase noise introduced by the Raman lasers. In addition, the difference in the measurements performed in configurations C and F (Fig. 1) efficiently rejects long-term systematic drifts which do not depend on the distribution of the source masses. They include wavefront distortions, magnetic field perturbations, Coriolis acceleration and light shift.

Controlling sources of statistical noise and systematic phase shifts in the gravity gradient measurement is crucial for the experiment[28]. Inertial effects induced by the Earth's rotation are an important source of both noise and systematic errors. Gravity gradient measurements are indeed sensitive to the Coriolis acceleration when differential velocities along the east–west direction are present between the two atomic samples and when the launch direction itself is not stable. Precise control of the effects of the Coriolis acceleration is achieved by applying a uniform rotation rate to the retroreflecting Raman mirror by means of PZT actuators during the atom interferometry sequence[23,24]. The optimal rotation rate, counteracting the local projection of the Earth's rotation rate on the horizontal plane, is determined to better than 2 μrad s^{-1} after maximizing the interferometers' contrast and minimizing the noise in the differential phase measurements. Under such conditions, the residual shift amounts to $1.5\Delta\theta$, where $\Delta\theta$ is the variation in the launch direction expressed in radians when switching from configuration C to configuration F, and vice versa. As a consequence, a G measurement down to the 100 p.p.m. level requires control of the launch direction to better than 37 μrad. We can control $\Delta\theta$ to better than 8 μrad. The velocity distribution of the atoms due to the finite temperature of the sample (~4 μK) also contributes a systematic shift in the G measurement. In our experiment, the Raman mirror tilt is controlled to better than 100 nrad, bringing the corresponding error in the G measurement to negligible levels. Magnetic fields can perturb atom interferometry measurements both through their mechanical action on the atomic trajectories and via the second-order Zeeman effect. A 10^3 shielding factor to external magnetic fields is achieved in the interferometer volume by passively isolating the vertical tube through a system of two cylindrical μ-metal layers. In addition, systematic shifts that slowly change over a typical timescale of a few seconds and that do not depend on the direction of the effective wavevector, k, are efficiently cancelled by reversing k with a frequency shift of the Raman lasers at each measurement cycle and taking the average phase shift[18]. They include the second-order Zeeman effect as well as the first-order light shift. Finally, the main parameters of the experiment have been actively stabilized to optimize the sensitivity and the long-term stability of gravity gradient measurements: the optical intensities of the cooling, Raman and probe laser beams are regulated by acting on the radio-frequency power driving the acousto-optical modulators, and the Raman mirror tilt is controlled by a piezo tip–tilt system. In this way, we obtain a sensitivity to differential accelerations of $3 \times 10^{-9} g$ for an integration time of 1 s, in agreement with the calculated quantum projection noise limit for 2×10^5 atoms[28].

Monte Carlo simulation and measurement systematics. Gravity gradient measurements are compared with a numerical simulation that evaluates the gravitational potential produced by a given configuration of the source masses, implements the calculation of the phase shift experienced by single atoms at the two interferometers, and finally runs a Monte Carlo simulation on the atomic trajectories by varying the initial position and velocity out of the density and velocity distribution of the atomic cloud that best fits the experimental density profiles. The gravitational potential generated by the source masses is computed analytically using a multipole expansion. The formula has been verified by calculating the potential of a single cylindrical mass along the classical path of the atoms during the interferometric sequence and comparing it against the numerical evaluation of the exact solution. The multipole expansion has been found to introduce negligible error while drastically reducing the computation time. The phase shift ϕ for an atom is obtained as $1/\hbar$ times the classical action, evaluated along the integration paths given by the solution of the Lagrange equation (equation (1)). We use a perturbative method[29] separating the Lagrangian L into $L_0 = p^2/2m - mgz$ (m is the atomic mass) and L_1, which accounts for the contributions of both the Earth's gravity gradient and the source mass. The phase shift due to L_1 is then calculated by integration over the unperturbed path given by the solution of the Lagrange equation for L_0. This approximation gives a negligible error in the determination of G because the ratio $|L_1/L_0|$ computed along the unperturbed paths is less than 10^{-7} in our experiment. The validity of the perturbative approach was additionally checked by comparison with the exact solution for the case of a uniform gradient[30]. The interaction processes between the atom and the Raman laser beams during the interferometry sequence are modelled in the code to account for the finite duration of the Raman pulses. At detection, we evaluate the convolution effects due to both the finite width of the detection beams and the finite bandwidth of the photodiodes collecting the fluorescence light in the $F = 1$ and $F = 2$ channels. Crucial for the Monte Carlo simulation is the spatial and momentum distribution used to describe the atomic sample and its evolution through the velocity selection pulses and the atom interferometer pulse sequence. Different techniques have been used to characterize the density and velocity profile of the atomic clouds. In particular, time-of-flight measurements provide the required information along the vertical direction, while the profiles in the horizontal plane are measured both by performing CCD imaging and by scanning the atomic cloud using the vertical Raman laser beams with a reduced diameter of 4 mm. The experimental data show that the spatial and velocity profiles of atomic clouds at launch, right after release from the magneto-optical trap, are well described by Gaussian distributions with spatial deviation ~3 mm and velocity deviation ~16 mm s^{-1} for the lower cloud and ~22 mm s^{-1} for the upper cloud. The velocity selection during interferometry reduces the vertical velocity deviation to ~3 mm s^{-1} and the horizontal velocity deviation to ~6 mm s^{-1}. Average transverse velocities are less than 1.5 mm s^{-1}.

The Monte Carlo simulation is used not only to evaluate the results of our measurements, but also to determine the sensitivity of the final phase angle to the relevant parameters and to estimate the uncertainty in the G measurement. Derivatives are computed with respect to 28 parameters, including the positions of the source masses; the positions and velocities of the upper and lower clouds and their times of flight at detection; the Gaussian widths of the detection lasers and their positions with respect to the atomic clouds; the interferometry sequence duration, $2T$; the effective wavevector, k, of the Raman lasers; and the gravitational acceleration and its gradient. From an analysis of the results, the spread of the atomic sample along the vertical direction appears to be responsible for the dominant contribution, requiring a measurement of the Gaussian width for both the upper and lower clouds to better than 0.1 mm. The determination of the atomic trajectories requires a good knowledge of the local gravitational acceleration and of the gravity gradient. Gravitational acceleration was measured by using an absolute gravimeter to an uncertainty that contributes a negligible error to the G measurement. A recent measurement of the gravity gradient performed with our atom interferometer gave 3.11×10^{-6} s^{-2}, with an effect of less than 2 p.p.m. on the determination of the Newtonian gravitational constant. The G measurement has been corrected for the effect of the finite density of the volume of air that replaces the source masses when they are moved from configuration C to configuration F, and vice versa. The corresponding correction to the G value amounts to $+60 \pm 6$ p.p.m. Finally, a correction has been applied to account for the non-uniform density of the tungsten cylinders[20];

©2014 Macmillan Publishers Limited. All rights reserved

the axial density inhomogeneity of the cylinders gives the main contribution, corresponding to a correction to the G value of $+90 \pm 20$ p.p.m.

The uncertainty budget of our measurement of the Newtonian gravitational constant is detailed in Table 1. Our result is reported in Fig. 3 and compared with the CODATA value[3] and recent determinations for G (refs 31–43).

30. Antoine, C. & Bordé, C. Exact phase shifts for atom interferometry. *Phys. Lett. A* **306,** 277–284 (2003).
31. Luther, G. G. & Towler, W. R. Redetermination of the Newtonian gravitational constant *G*. *Phys. Rev. Lett.* **48,** 121–123 (1982).
32. Karagioz, O. & Izmailov, V. Measurement of the gravitational constant with a torsion balance. *Meas. Tech.* **39,** 979–987 (1996).
33. Bagley, C. H. & Luther, G. G. Preliminary results of a determination of the Newtonian constant of gravitation: a test of the Kuroda hypothesis. *Phys. Rev. Lett.* **78,** 3047–3050 (1997).
34. Gundlach, J. H. & Merkowitz, S. M. Measurement of Newton's constant using a torsion balance with angular acceleration feedback. *Phys. Rev. Lett.* **85,** 2869–2872 (2000).
35. Quinn, T. J., Speake, C. C., Richman, S. J., Davis, R. S. & Picard, A. A new determination of *G* using two methods. *Phys. Rev. Lett.* **87,** 111101 (2001).
36. Kleinevoss, U. *Bestimmung der Newtonschen Gravitationskonstanten G*. PhD thesis, Univ. Wuppertal (2002).
37. Armstrong, T. R. & Fitzgerald, M. P. New measurements of *G* using the measurement standards laboratory torsion balance. *Phys. Rev. Lett.* **91,** 201101 (2003).
38. Hu, Z.-K., Guo, J.-Q. & Luo, J. Correction of source mass effects in the HUST-99 measurement of *G*. *Phys. Rev. D* **71,** 127505 (2005).
39. Schlamminger, S. *et al.* Measurement of Newton's gravitational constant. *Phys. Rev. D* **74,** 082001 (2006).
40. Luo, J. *et al.* Determination of the Newtonian gravitational constant *G* with time-of-swing method. *Phys. Rev. Lett.* **102,** 240801 (2009).
41. Tu, L.-C. *et al.* New determination of the gravitational constant *G* with time-of-swing method. *Phys. Rev. D* **82,** 022001 (2010).
42. Parks, H. V. & Faller, J. E. Simple pendulum determination of the gravitational constant. *Phys. Rev. Lett.* **105,** 110801 (2010).
43. Quinn, T., Parks, H., Speake, C. & Davis, R. Improved determination of *G* using two methods. *Phys. Rev. Lett.* **111,** 101102 (2013).

©2014 **Macmillan Publishers Limited. All rights reserved**

LETTER

doi:10.1038/nature13434

Chirality-specific growth of single-walled carbon nanotubes on solid alloy catalysts

Feng Yang[1], Xiao Wang[1], Daqi Zhang[1], Juan Yang[1], Da Luo[1], Ziwei Xu[2], Jiake Wei[3], Jian-Qiang Wang[4], Zhi Xu[3], Fei Peng[1], Xuemei Li[5], Ruoming Li[1], Yilun Li[1], Meihui Li[1], Xuedong Bai[3], Feng Ding[2] & Yan Li[1]

Carbon nanotubes have many material properties that make them attractive for applications[1,2]. In the context of nanoelectronics[3], interest has focused on single-walled carbon nanotubes[4] (SWNTs) because slight changes in tube diameter and wrapping angle, defined by the chirality indices (*n*, *m*), will shift their electrical conductivity from one characteristic of a metallic state to one characteristic of a semiconducting state, and will also change the bandgap. However, this structure–function relationship can be fully exploited only with structurally pure SWNTs. Solution-based separation methods[5] yield tubes within a narrow structure range, but the ultimate goal of producing just one type of SWNT by controlling its structure during growth has proved to be a considerable challenge over the last two decades[6–9]. Such efforts aim to optimize the composition[6,10–17] or shape[18–21] of the catalyst particles that are used in the chemical vapour deposition synthesis process to decompose the carbon feedstock and influence SWNT nucleation and growth[22–25]. This approach resulted in the highest reported proportion, 55 per cent, of single-chirality SWNTs in an as-grown sample[11]. Here we show that SWNTs of a single chirality, (12, 6), can be produced directly with an abundance higher than 92 per cent when using tungsten-based bimetallic alloy nanocrystals as catalysts. These, unlike other catalysts used so far, have such high melting points that they maintain their crystalline structure during the chemical vapour deposition process. This feature seems crucial because experiment and simulation both suggest that the highly selective growth of (12, 6) SWNTs is the result of a good structural match between the carbon atom arrangement around the nanotube circumference and the arrangement of the catalytically active atoms in one of the planes of the nanocrystal catalyst. We anticipate that using high-melting-point alloy nanocrystals with optimized structures as catalysts paves the way for total chirality control in SWNT growth and will thus promote the development of SWNT applications.

The basic idea underpinning our study is sketched in Fig. 1: we use molecular clusters to prepare alloy nanoparticles with specific structural features that can then direct the structure-specific growth of SWNTs. Tungsten-based alloys normally form at extremely high temperature, but W–Co alloy nanoparticles can be obtained under moderate conditions when using as precursor the molecular cluster $Na_{15}[Na_3 \subset \{Co(H_2O)_4\}_6\{WO(H_2O)\}_3(P_2W_{12}O_{48})_3]\cdot nH_2O$ (ref. 26; denoted as $\{W_{39}Co_6O_x\}$), where '⊂' indicates that the three Na atoms preceding the symbol are encapsulated by the units listed after the symbol. The resultant alloy nanoparticles were then used to catalyse SWNT growth on SiO_2/Si substrates via ethanol chemical vapour deposition (CVD).

Micro-Raman measurements (Extended Data Fig. 1) on SWNTs grown at 1,030 °C (Fig. 2c) show radial breathing mode (RBM) peaks intensely concentrated at ~197 cm^{-1} (Fig. 2a, b and Supplementary Video 1). Surface-enhanced Raman scattering measurements, which provide Raman spectra of individual SWNTs regardless of whether they are resonant with the excitation laser or not[27], further confirm the domination of the RBM peaks at ~197 cm^{-1} (Extended Data Fig. 2). Tube diameter measurements also testify to the uniformity of the SWNT sample produced: the height profile of the atomic force microscope (AFM) image indicates a uniform diameter (d_t) of ~1.28 nm (Extended Data Fig. 3), which closely matches the d_t value of 1.24 nm calculated from the RBM frequency (Supplementary equation (1)). The corresponding tangential vibration (G mode) bands at ~1,580 cm^{-1} exhibit typical Breit–Wigner–Fano line shapes (Fig. 2b and Supplementary Fig. 3), a feature typical of metallic SWNTs[28]. A chirality of (12, 6) was assigned by comparing with Kataura plot[28] (Extended Data Fig. 4 and Supplementary Fig. 2). This assignment was further verified by the electron diffraction measurements on isolated tubes grown across the slits of silicon substrates and with RBMs at ~197 cm^{-1}, which all show typical electron diffraction patterns indexed to (12, 6) (Fig. 2d, Extended Data Fig. 5 and Supplementary Fig. 4).

The abundance of (12, 6) tubes in the samples grown at 1,030 °C was quantified by three methods. The statistics of the RBMs in the micro-Raman spectra, and counting for the coverage of excitation, gives an estimate of 94.4% based on ~3,300 detected tubes (Fig. 2f, Supplementary Tables 2 and 3). The method reported in ref. 29, which uses AFM to count all tubes and Raman spectroscopy to identify (12, 6) tubes, gives a fraction of 94.9% (Extended Data Fig. 6). And the ultraviolet/visible/near-infrared (UV–Vis–NIR) absorption spectrum[30] of aqueous dispersions collected from 42 SWNT samples (Extended Data Fig. 7) indicates that the content of (12, 6) tubes is 92.5% (Fig. 2e and Supplementary Table 4). This analysis demonstrates direct growth of SWNTs with a single dominating chirality in a proportion greater than 92%.

To explore what enables this unprecedented chirality-selective SWNT growth, we carefully studied the structure of our catalyst nanoparticles. The extended X-ray absorption fine structure data show that W and Co are alloyed at 1,030 °C, with a W–Co bond length of 2.2 Å (Fig. 3a). The X-ray diffraction (XRD) patterns of these catalysts, supported on SiO_2,

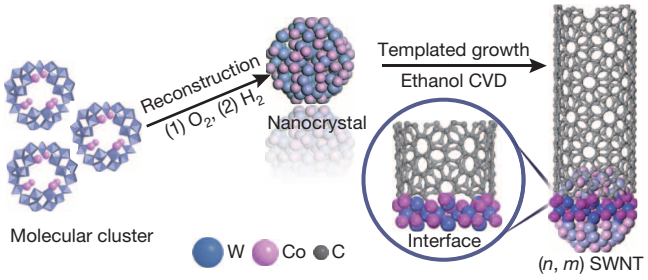


Figure 1 | Preparation of the W–Co nanocrystal catalyst and the templated growth of a SWNT with specified (*n*, *m*).

[1]Beijing National Laboratory for Molecular Science, Key Laboratory for the Physics and Chemistry of Nanodevices, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China. [2]Institute of Textiles and Clothing, Hong Kong Polytechnic University, Kowloon, Hong Kong, China. [3]Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. [4]Shanghai Synchrotron Radiation Facility, Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201204, China. [5]Electron Microscopy Laboratory, Peking University, Beijing 100871, China.

©2014 Macmillan Publishers Limited. All rights reserved

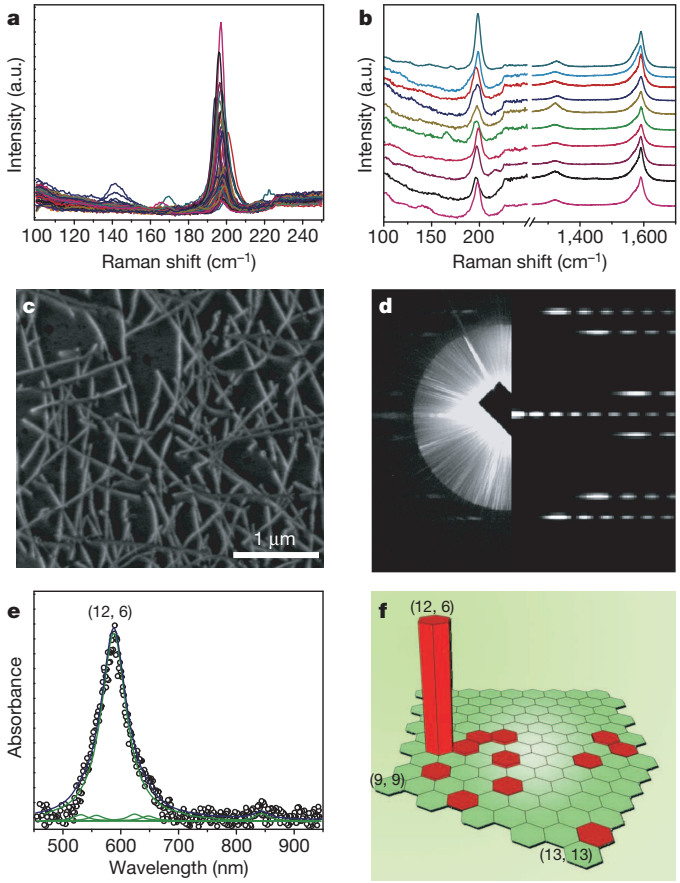


Figure 2 | Characterizations of SWNTs grown at 1,030 °C. **a**, RBM-region Raman spectra obtained with a 1-μm laser spot. a.u., arbitrary units. **b**, Raman spectra taken with a 5-μm laser spot (excitation wavelength, 633 nm). **c**, Scanning electron microscope (SEM) image. **d**, Electron diffraction pattern companied by the standard pattern of a (12, 6) tube. **e**, UV–Vis–NIR absorption spectrum (background-subtracted) (open circles) of an aqueous dispersion collected from 42 samples, deconvolution curves (green solid line) and fitting spectrum (blue solid line). **f**, Relative abundances of various chiralities from Raman measurements from ~3,300 nanotubes denoted in the SWNT chiral map.

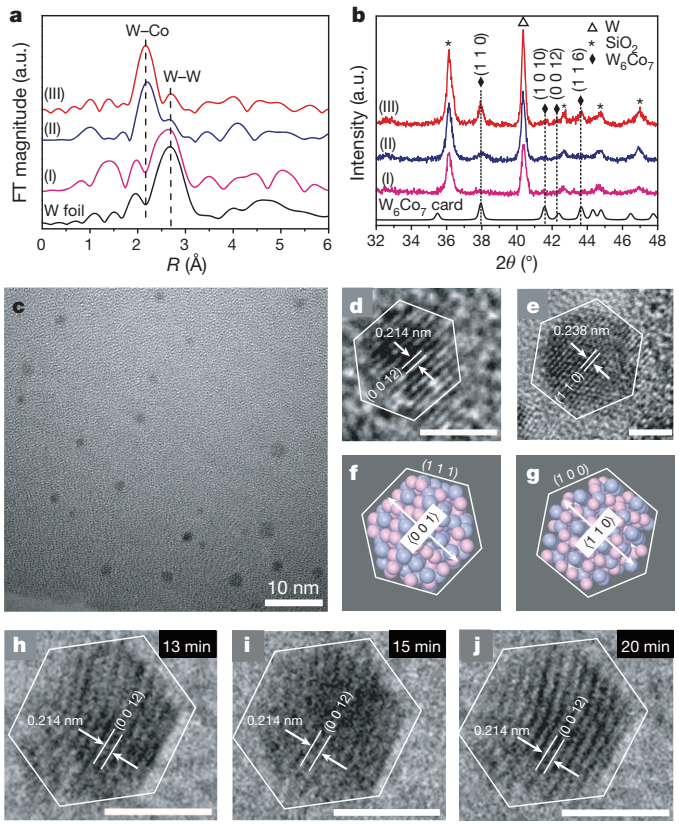


Figure 3 | Characterizations of catalyst structure. **a**, **b**, Fourier transforms (FTs) of the *L*-edge extended X-ray absorption fine structure data of W (without phase shift) (**a**) and XRD patterns (**b**) obtained on the fresh reduced catalysts at 950 (line I), 1,000 (line II) and 1,030 °C (line III). Here *R* is the distance between two adjacent atoms and 2θ is the X-ray scattering angle. **c–e**, TEM and HRTEM images of catalyst particles prepared at 1,030 °C. **f**, **g**, Polyhedral atomic models of W_6Co_7 alloy nanocrystals (pink, Co; blue, W). **h–j**, *In situ* HRTEM images of W–Co catalyst particles taken at 1,100 °C in vacuum. Scale bars in **d**, **e**, **h–j** are 2 nm.

present diffraction peaks corresponding to the (1 1 0), (1 1 6), (0 0 12) and (1 0 10) planes of W_6Co_7 alloy (Fig. 3b). A transmission electron microscope (TEM) image of the particles is shown in Fig. 3c, and high-resolution TEM (HRTEM) images that exhibit lattice fringes in good accordance with the plane distances of W_6Co_7 are shown in Fig. 3d, e (the orientations of the corresponding polyhydric nanocrystals are schematically illustrated in Fig. 3f, g).

Only diffraction peaks corresponding to W_6Co_7 alloy, W and SiO_2 appear in the XRD patterns of the catalysts. Control experiments show no SWNT growth when using pure W as catalyst under typical CVD conditions. XRD measurements further show that W_6Co_7 alloy is still present after the ethanol CVD process (Extended Data Fig. 8). These observations indicate that the W_6Co_7 alloy nanoparticles are responsible for catalysing the growth of (12, 6) SWNTs.

We note that bulk W_6Co_7 alloy has a high melting point of ~2,400 °C, and *in situ* TEM measurements reveal that the W_6Co_7 nanoparticles maintain their crystalline structure with well-faceted shapes when heated in vacuum to 1,100 °C (Fig. 3h–j). This suggests that the catalyst nanoparticles are still solid during the CVD process at 1,030 °C.

The interface between a catalyst particle and an attached SWNT (seen in the SEM image in Fig. 4a) was probed using *ex situ* HRTEM (Supplementary Fig. 6). The suspended tube with RBM at ~197 cm^{-1} shows a typical electron diffraction pattern indexed to (12, 6) (Fig. 4b, c). The HRTEM image in Fig. 4d shows that the interplane distance of the catalyst along the axis of the (12, 6) SWNT is in good accordance with that of the (0 0 12) planes in W_6Co_7 alloy. The corresponding fast Fourier transform pattern (Fig. 4e) further verifies that the (12, 6) tube axis is perpendicular to (0 0 12) planes. Figure 4f shows a SWNT cap nucleating around the (0 0 12) plane of a W_6Co_7 nanoparticle.

When using density functional theory simulations to fit different SWNTs to the (0 0 12) plane of a W_6Co_7 alloy particle, we found a perfect geometrical match for the (12, 6) tube and a poor match for all other tubes examined (Extended Data Fig. 9). The simulations also revealed that the lower uniformity in the arrangement of the catalytically active Co atoms in W_6Co_7 alloy, relative to that in pure Co particles, results in a more specific structural match between certain SWNTs and the catalyst nanoparticles. We suggest that the structural match between catalyst and nanotube might be an essential factor in the chirality-specific growth we have achieved. This is further supported by control experiments that use either a catalyst prepared at 1,030 °C to perform CVD nanotube growth at reduced temperatures of 1,000 and 950 °C, or catalysts prepared at 1,000 and 950 °C (at which W and Co are not alloyed): the former experiment produced a remarkably higher abundance of (12, 6) SWNTs (Extended Data Fig. 10), indicating that the alloy catalyst is critical for the growth of (12, 6) tubes. Initial experiments have shown that other SWNT types, for example (16, 0) and (14, 4), can be produced in a chirality-selective manner when using W_6Co_7 nanoparticles prepared under different conditions (Supplementary Figs 7 and 8). Our strategy was further proved valid using W–Fe and W–Ni bimetallic catalysts (Supplementary Figs 9 and 10).

Our observations demonstrate that W-based alloy nanoparticles are effective catalysts for chirality-specific SWNT growth, and suggest that

©2014 Macmillan Publishers Limited. All rights reserved

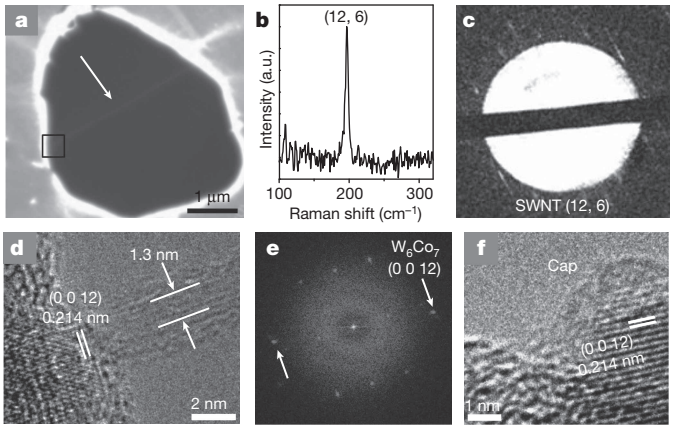


Figure 4 | Characterizations of SWNTs nucleated on catalyst nanoparticles. **a**, SEM image of an individual suspended SWNT (indicated by the arrow). **b**, **c**, Raman spectrum (**b**) and corresponding electron diffraction pattern (**c**) of this SWNT. **d**, HRTEM image of the area marked in **a** showing the axis of this SWNT perpendicular to the (0 0 12) fringes of W_6Co_7 nanoparticles. **e**, Corresponding fast Fourier transforms of this W_6Co_7 nanoparticle, spots corresponding to (0 0 12) planes are indicated using arrows. **f**, HRTEM image of a SWNT cap nucleated around the (0 0 12) plane of a W_6Co_7 nanocrystal.

their very high melting points and unique atomic arrangements are the two crucial factors enabling selective growth. We believe that the growth of SWNTs with chirality purity greater than 99% is possible with further careful manipulation of the composition, structure and size of the alloy catalyst nanoparticles and optimization of the CVD conditions.

METHODS SUMMARY

The molecular cluster $Na_{15}[Na_3\subset\{Co(H_2O)_4\}_6\{WO(H_2O)\}_3(P_2W_{12}O_{48})_3]\cdot nH_2O$ (ref. 26) was used as the catalyst precursor. The molecular clusters were deposited on silicon wafers, calcined in air and then reduced with H_2 through temperature-programmed reduction processes to obtain various catalysts. Then CVD growth of carbon nanotubes was performed at different temperatures with ethanol as the carbon feedstock. SEM, AFM and micro-Raman spectroscopy were used to characterize the as-grown samples. The chirality assignment was performed on the basis of Raman spectroscopy and electron diffraction patterns. The chirality distribution was determined by three methods. The as-grown sample was measured by counting the RBM peaks. Then SWNTs were dispersed in aqueous solution and the UV–Vis–NIR absorption was performed to determine the chirality composition. Then this aqueous dispersion of SWNTs was dropped onto a silicon wafer and the content of (12, 6) tubes was obtained by counting the total number of SWNTs using AFM and counting the number of (12, 6) tubes using Raman mapping[29]. The catalyst structure was characterized by X-ray absorption spectroscopy, XRD and HRTEM. The *in situ* HRTEM of the catalyst nanoparticles was performed at 1,100 °C in vacuum. The self-consistent density functional theory simulations were carried out using the Vienna Ab-initio Simulation Package. More details are described in the Supplementary Information.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 February 2013; accepted 17 April 2014.

1. Shulaker, M. M. *et al.* Carbon nanotube computer. *Nature* **501,** 526–530 (2013).
2. De Volder, M. F. L., Tawfick, S. H., Baughman, R. H. & Hart, A. J. Carbon Nanotubes: present and future commercial applications. *Science* **339,** 535–539 (2013).
3. Avouris, P., Chen, Z. & Perebeinos, V. Carbon-based electronics. *Nature Nanotechnol.* **2,** 605–615 (2007).
4. Iijima, S. & Ichihashi, T. Single-shell carbon nanotubes of 1-nm diameter. *Nature* **363,** 603–605 (1993).
5. Hersam, M. C. Progress towards monodisperse single-walled carbon nanotubes. *Nature Nanotechnol.* **3,** 387–394 (2008).
6. Bachilo, S. M. *et al.* Narrow (*n,m*)-distribution of single-walled carbon nanotubes grown using a solid supported catalyst. *J. Am. Chem. Soc.* **125,** 11186–11187 (2003).
7. Ding, F., Harutyunyan, A. R. & Yakobson, B. I. Dislocation theory of chirality-controlled nanotube growth. *Proc. Natl Acad. Sci. USA* **106,** 2506–2509 (2009).
8. Liu, J. *et al.* Chirality-controlled synthesis of single-wall carbon nanotubes using vapour-phase epitaxy. *Nature Commun.* **3,** 1199 (2012).
9. Chen, Y. *et al.* Helicity-dependent single-walled carbon nanotube alignment on graphite for helical angle and handedness recognition. *Nature Commun.* **4,** 2205 (2013).
10. An, L., Owens, J. M., McNeil, L. E. & Liu, J. Synthesis of nearly uniform single-walled carbon nanotubes using identical metal-containing molecular nanoclusters as catalysts. *J. Am. Chem. Soc.* **124,** 13688–13689 (2002).
11. Lolli, G. *et al.* Tailoring (*n*, *m*) structure of single-walled carbon nanotubes by modifying reaction conditions and the nature of the support of CoMo catalysts. *J. Phys. Chem. B* **110,** 2108–2115 (2006).
12. Zhou, W. *et al.* Copper catalyzing growth of single-walled carbon nanotubes on substrates. *Nano Lett.* **6,** 2987–2990 (2006).
13. Li, X. *et al.* Selective synthesis combined with chemical separation of single-walled carbon nanotubes for chirality selection. *J. Am. Chem. Soc.* **129,** 15770–15771 (2007).
14. Chiang, W. H. & Sankaran, R. M. Linking catalyst composition to chirality distributions of as-grown single-walled carbon nanotubes by tuning Ni_xFe_{1-x} nanoparticles. *Nature Mater.* **8,** 882–886 (2009).
15. Wang, H. *et al.* Chiral-selective $CoSO_4/SiO_2$ catalyst for (9,8) single-walled carbon nanotube growth. *ACS Nano* **7,** 614–626 (2013).
16. Liu, B., Ren, W., Li, S., Liu, C. & Cheng, H.-M. High temperature selective growth of single-walled carbon nanotubes with a narrow chirality distribution from a CoPt bimetallic catalyst. *Chem. Commun.* **48,** 2409–2411 (2012).
17. Peng, F. *et al.* Diameter-controlled growth of aligned single-walled carbon nanotubes on quartz using molecular nanoclusters as catalyst precursors. *Chin. Sci. Bull.* **58,** 433–439 (2013).
18. Reich, S., Li, L. & Robertson, J. Control the chirality of carbon nanotubes by epitaxial growth. *Chem. Phys. Lett.* **421,** 469–472 (2006).
19. Harutyunyan, A. R. *et al.* Preferential growth of single-walled carbon nanotubes with metallic conductivity. *Science* **326,** 116–120 (2009).
20. Koziol, K. K. K., Ducati, C. & Windle, A. H. Carbon nanotubes with catalyst controlled chiral angle. *Chem. Mater.* **22,** 4904–4911 (2010).
21. He, M. *et al.* Chiral-selective growth of single-walled carbon nanotubes on lattice-mismatched epitaxial cobalt nanoparticles. *Sci. Rep.* **3,** 1460 (2013).
22. Li, Y. *et al.* How catalysts affect the growth of single-walled carbon nanotubes on substrates. *Adv. Mater.* **22,** 1508–1515 (2010).
23. Shibuta, Y. & Maruyama, S. Molecular dynamics simulation of formation process of single-walled carbon nanotubes by CCVD method. *Chem. Phys. Lett.* **382,** 381–386 (2003).
24. Raty, J.-Y., Gygi, F. & Galli, G. Growth of carbon nanotubes on metal nanoparticles: a microscopic mechanism from ab initio molecular dynamics simulations. *Phys. Rev. Lett.* **95,** 096103 (2005).
25. Hofmann, S. *et al.* In situ observations of catalyst dynamics during surface-bound carbon nanotube nucleation. *Nano Lett.* **7,** 602–608 (2007).
26. Yao, S., Zhang, Z., Li, Y. & Wang, E. Two hexa-TM-containing (TM = Co^{2+} and Ni^{2+}) $\{P_2W_{12}\}$-based trimeric tungstophosphates. *Dalton Trans.* **39,** 3884–3889 (2010).
27. Chu, H. *et al.* Decoration of gold nanoparticles on surface-grown single-walled carbon nanotubes for detection of every nanotube by surface-enhanced Raman spectroscopy. *J. Am. Chem. Soc.* **131,** 14310–14316 (2009).
28. Dresselhaus, M. S., Dresselhaus, G., Saito, R. & Jorio, A. Raman spectroscopy of carbon nanotubes. *Phys. Rep.* **409,** 47–99 (2005).
29. Naumov, A. V. *et al.* Quantifying the semiconducting fraction in single-walled carbon nanotube samples through comparative atomic force and photoluminescence microscopies. *Nano Lett.* **9,** 3203–3208 (2009).
30. Strano, M. S. Electronic structure control of single-walled carbon nanotube functionalization. *Science* **301,** 1519–1522 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Iijima, S. Maruyama and J. Liu for advice on the manuscript; K. Lu, Y. Wang, E. Wang and S. Yao for discussions on catalyst preparation and properties; D. P. Nackashi, J. Ju, Y. Li and Protochips Inc. for *in situ* TEM measurements; H. Xu, H. Yang, K. Jiang and S. Wang for Raman measurements. This research was financially supported by MoST (project 2011CB933003), the NSFC of China (projects 21125103, 91333105, 21321001, 11179011 and 21273189) and a Hong Kong GRF research grant (G-YX4Q).

Author Contributions Yan Li and F.Y. had the idea for and designed the experiments, analysed the results and wrote the manuscript. F.Y., D.Z., J.Y., D.L., J.W., J.-Q.W., Zhi Xu, X.L., F.P., Yilun Li, R.L., M.L. and X.B. performed the experiments and analysed the results. X.W., Ziwei Xu and F.D. contributed to theoretical calculations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Yan Li (yanli@pku.edu.cn).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13456

South Greenland ice-sheet collapse during Marine Isotope Stage 11

Alberto V. Reyes[1,2]†, Anders E. Carlson[1,3], Brian L. Beard[1], Robert G. Hatfield[3], Joseph S. Stoner[3], Kelsey Winsor[1], Bethany Welke[1]† & David J. Ullman[1,3]

Varying levels of boreal summer insolation and associated Earth system feedbacks led to differing climate and ice-sheet states during late-Quaternary interglaciations. In particular, Marine Isotope Stage (MIS) 11 was an exceptionally long interglaciation and potentially had a global mean sea level 6 to 13 metres above the present level around 410,000 to 400,000 years ago[1,2], implying substantial mass loss from the Greenland ice sheet (GIS). There are, however, no model simulations and only limited proxy data[3,4] to constrain the magnitude of the GIS response to climate change during this 'super interglacial'[5], thus confounding efforts to assess climate/ice-sheet threshold behaviour[6,7] and associated sea-level rise[1,2]. Here we show that the south GIS was drastically smaller during MIS 11 than it is now, with only a small residual ice dome over southernmost Greenland. We use the strontium–neodymium–lead isotopic composition of proglacial sediment discharged from south Greenland to constrain the provenance of terrigenous silt deposited on the Eirik Drift, a sedimentary deposit off the south Greenland margin. We identify a major reduction in sediment input derived from south Greenland's Precambrian bedrock terranes, probably reflecting the cessation of subglacial erosion and sediment transport[8] as a result of near-complete deglaciation of south Greenland. Comparison with ice-sheet configurations from numerical models[7,9–12] suggests that the GIS lost about 4.5 to 6 metres of sea-level-equivalent volume during MIS 11. This is evidence for late-Quaternary GIS collapse after it crossed a climate/ice-sheet stability threshold that may have been no more than several degrees above pre-industrial temperatures[6,7].

Little information is available on the magnitude of the retreat of the GIS or the Antarctic ice sheet (AIS) during MIS 11, thus hindering efforts to understand potentially nonlinear responses of Earth's ice sheets to protracted warm intervals. AIS retreat could explain all of the proposed +6–13 m MIS 11 sea-level range[1,13]. Alternatively, near-complete GIS deglaciation would account for the lower-end estimate of the sea-level highstand[14], with some AIS contribution required for a highstand greater than +7 m. Yet another possibility is that the GIS and AIS experienced minimal retreat during MIS 11, consistent with reconstructions suggesting that MIS 11 sea level may not have been appreciably different from today[2]. Thus, constraints on the MIS 11 response of the GIS can shed considerable light on uncertain palaeo-sea-level records, sparse palaeoclimate proxy data[5] and the possibility that the modelled temperature threshold for GIS collapse[6,7] may have been exceeded during MIS 11. In the absence of direct geomorphic evidence for the magnitude of ice-sheet retreat during past interglaciations, we turn to geochemical proxy evidence to assess the south GIS response to MIS 11 climate change and, in turn, estimate the GIS contribution to MIS 11 sea level.

We use the sedimentary record of Eirik Drift core MD99-2227 (Figs 1 and 2) to infer GIS extent on south Greenland, where many models suggest interglacial ice-sheet mass loss will be most pronounced[9–12]. The Eirik Drift receives terrigenous sediment from subglacial erosion of south Greenland's Precambrian bedrock (the Ketilidian Mobile Belt (KMB), Archaean Block (AB) and Nagssugtoqidian Mobile Belt (NMB)) and Palaeogene volcanics that outcrop in east Greenland and Iceland (Fig. 1). These sediments are transported to the core site in the Western Boundary Undercurrent[15–17] (Fig. 1 and Methods), and magnetic and geochemical properties of Holocene Eirik Drift sediments document the close coupling between GIS subglacial erosion, sediment transport and terrigenous silt deposition at MD99-2227[16–18]. Other potential sources of terrigenous silt at Eirik Drift, including northeast Greenland, Scotland, Scandinavia and re-transported silt in Arctic sea ice, are likely to be negligible (Methods).

We estimate the provenance of the carbonate-free detrital silt fraction in Eirik Drift sediments using Sr–Nd–Pb isotope ratios, which trace terrestrial silt sources because the unique isotope compositions of south Greenland's bedrock terranes reflect their differing ages and

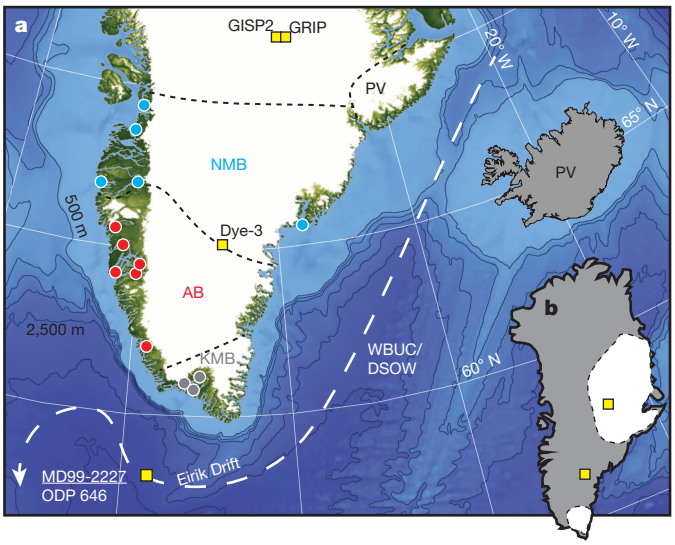


Figure 1 | Map of Greenland and other features mentioned in the text. **a**, Bedrock terrane boundaries (black dashed lines) and sampling sites for Greenland stream sediments (filled circles; multiple sites per symbol); KMB, Ketilidian Mobile Belt (grey); AB, Archaean Block (red); NMB, Nagssugtoqidian Mobile Belt (blue); PV, Palaeogene volcanics. Yellow squares mark locations of marine sediment core MD99-2227 and Dye-3 and Summit (GISP2, GRIP) ice cores. Dashed white lines denote modern deep-water circulation features that are thought to have been active during past interglaciations[15–21]; DSOW, Denmark Strait Overflow Water; WBUC, Western Boundary Undercurrent. The bathymetric contour interval is 500 m. **b**, Inset map. White polygons show the potential configuration of the MIS 11 Greenland ice sheet, which is similar to modelled ice limits representing ~6 m of sea-level-equivalent mass loss[7] (Methods). Yellow squares are ice-core sites shown in **a**.

[1]Department of Geoscience, University of Wisconsin-Madison, 1215 West Dayton Street, Madison, Wisconsin 53706, USA. [2]School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, Elmwood Avenue, Belfast BT7 1NN, UK. [3]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, 104 CEOAS Administration Building, Corvallis, Oregon 97331, USA. †Present addresses: Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada (A.V.R); Department of Geosciences, University of Arizona, 1040 East 4th Street, Tucson, Arizona 85721, USA (B.W.).

©2014 Macmillan Publishers Limited. All rights reserved

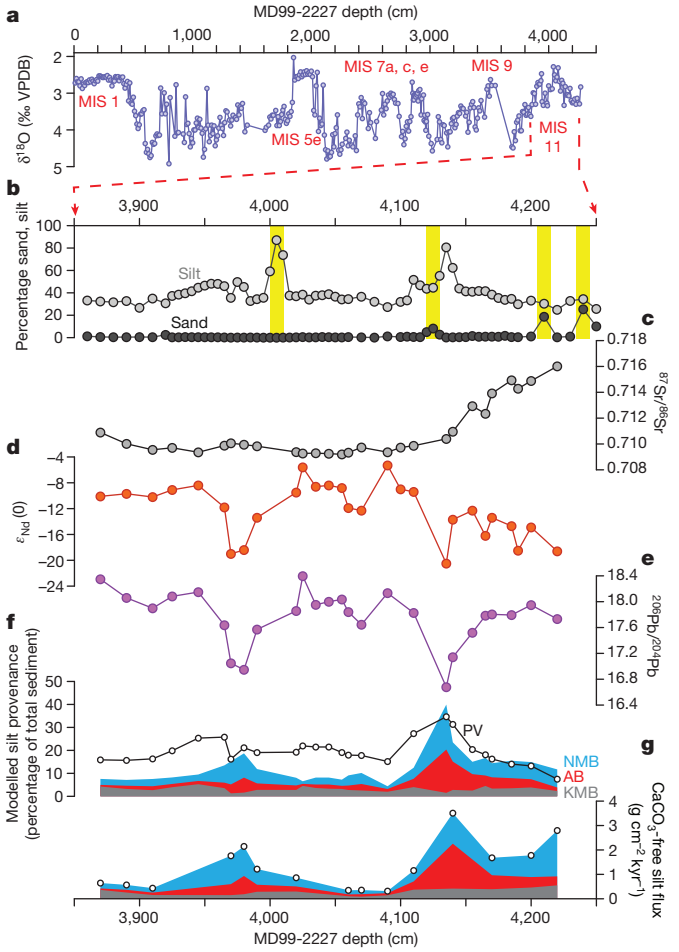


Figure 2 | MD99-2227 records on depth scale. a, Full $\delta^{18}O$ *Neogloboquadrina pachyderma* (sinistral) record[19]; interglacial MISs are indicated. $\delta^{18}O = 1{,}000 \times [(^{18}O/^{16}O)_{sample}/(^{18}O/^{16}O)_{VPDB} - 1]$; VPDB, Vienna PeeDee belemnite. **b**, Weight per cent silt and sand. Vertical yellow bars denote ice-rafted debris and turbidite layers[19] that are not included in silt provenance estimates. **c–e**, Isotopic composition, with analytical uncertainty smaller than symbol size: $^{87}Sr/^{86}Sr$ (**c**); ε_{Nd} (**d**; Methods); $^{206}Pb/^{204}Pb$ (**e**). $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$ follow a similar trend (Supplementary Table 1) and are not shown here. **f**, Median inferred silt provenance from isotope mixing model, expressed as percentage of total sediment. Provenance estimates do not sum to 100% because they do not include clay or sand size fractions. **g**, Estimated flux of $CaCO_3$-free silt from south Greenland Precambrian terranes (upper and lower uncertainty estimates in Extended Data Fig. 2).

tectonometamorphic histories[16] (Methods and Extended Data Fig. 1). Silt from KMB sources is readily distinguished from the AB and NMB by higher $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$ and ε_{Nd} values (Methods). The AB and NMB are more similar isotopically, although AB silts tend towards higher $^{87}Sr/^{86}Sr$, and lower $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$, than do NMB silts. Silts derived from Palaeogene volcanic rocks have very low $^{87}Sr/^{86}Sr$ and high ε_{Nd}, compared with the three Precambrian terranes.

The MD99-2227 $\delta^{18}O$ stratigraphy indicates the presence of MIS 11 sediments below ~3,950 cm core depth, and an age–depth model based on $\delta^{18}O$ and relative geomagnetic palaeointensity suggests that the base of MD99-2227 dates to late MIS 12[19] (Fig. 2a and Methods). The $^{87}Sr/^{86}Sr$ ratio decreases from ~0.717 to ~0.709 between depths of 4,240 and 4,090 cm (Fig. 2c). The $^{87}Sr/^{86}Sr$ ratio stays at or below ~0.709 until 3,990 cm, where it increases to ~0.711 near the end of MIS 11. Though the ε_{Nd} data are more variable (Fig. 2d), with two excursions to low ε_{Nd} at 4,135 and ~3,975 cm, ε_{Nd} follows a general increasing trend from −20.9 to −5.3 between 4,240 and 4,090 cm, followed by a return to lower ε_{Nd} at the end of MIS 11. The Pb isotope data mimic the ε_{Nd} trend through MIS 11 (Fig. 2e).

Because south Greenland Precambrian terranes have distinctly high $^{87}Sr/^{86}Sr$ and low ε_{Nd}, the Sr and Nd isotope data are consistent with an overall reduction in silt derived from south Greenland's Precambrian bedrock. Direct interpretation of the Pb data is complicated by anomalous KMB Pb isotope compositions relative to the AB and NMB (Extended Data Fig. 1). We use an isotope mixing model to estimate the proportion and flux of silt derived from each south Greenland terrane, based on the median solution from 10,000 model runs (Fig. 2f, g and Methods). The Palaeogene volcanic silt fraction of the total sediment increases from ~10% during late MIS 12 to 15–30% during MIS 11 (Fig. 2f), which probably reflects intensification of the Western Boundary Undercurrent during interglaciations[15–21]. Precambrian silt is initially 10–15% of the total sediment during MIS 12, with an increase to ~40% at 4,135 cm early in MIS 11 that is mainly driven by increases in the AB and NMB fractions. Precambrian silt decreases to 4–8% of the total sediment for the middle of MIS 11 (4,090–4,020 cm), and then increases again, to ~20% at 3,980 cm, with a subsequent return to ~5% at the end of MIS 11. Because inferred KMB contributions remain relatively constant during this interval, at <4% of total sediment, the decrease in Precambrian silt sedimentation above 4,110 cm is driven mainly by changes in silt sourced from the AB and NMB terranes.

Sediment flux estimates (Fig. 2g, Methods and Extended Data Fig. 2) corroborate the provenance interpretations based on the percentage of total sediment. Silt flux from south Greenland's Precambrian bedrock is 1.7–3.5 g cm^{-2} kyr^{-1} during late MIS 12 and into MIS 11, but decreases by an order of magnitude to ~0.3 g cm^{-2} kyr^{-1} during MIS 11. Flux from the KMB during MIS 11 is low, but does not decrease as precipitously as flux from the AB or NMB, both of which decrease to a flux of less than ~0.1 g cm^{-2} kyr^{-1}.

Comparing our terrane provenance record with similar estimates for the last interglaciation (LIG; 128–116 kyr ago) and the Holocene epoch[16], we find that the Precambrian silt contribution at the end of MIS 12 is similar to sedimentation during the penultimate and last deglaciations (Fig. 3b, c). Precambrian silt early in MIS 11 is largely derived from the NMB and AB, and probably reflects increased ablation and attendant discharge of silt from the south GIS, similar to early LIG and Holocene silt pulses[16]. The low in Precambrian silt during the MIS 11 $\delta^{18}O$ minimum (Fig. 3a) is unprecedented with respect to these later interglaciations (Fig. 3b). The near-complete loss of NMB- and AB-derived silt drives this minimum, with individual terrane contributions decreasing to <1% of total sediment. For comparison, the minimum NMB + AB fraction is 10–15% of the total sediment during the Holocene and LIG (Fig. 3b). Although it is not appropriate to compare sediment flux during the Holocene and Termination I owing to stretching in the upper part of MD99-2227[19] (Methods), terrigenous silt flux from south Greenland's Precambrian bedrock was over an order of magnitude higher in the LIG than during MIS 11 (Fig. 3c).

One explanation for the near absence of NMB and AB silt during the MIS 11 $\delta^{18}O$ minimum is that a change in ocean circulation reduced or stopped transport of NMB and AB sediment to Eirik Drift. However, the continued input of KMB and Palaeogene volcanic silt suggests that sediment transport pathways to Eirik Drift during MIS 11 were similar to other interglaciations[15,16,21]. Alternatively, ablation, meltwater production and subglacial erosion are high during ice-sheet retreat[8], but complete ice loss from a given terrane would lead to the cessation of subglacial erosion and meltwater transport of silt and, in turn, to decreased sediment flux to the ocean. We therefore interpret the near absence of NMB and AB silt during the MIS 11 $\delta^{18}O$ minimum as indicating the loss of ice from these terranes. Deglaciated terranes may supply minor amounts of silt even during periods of minimal or absent GIS coverage, owing to paraglacial processes that entrain and transport glacial sediment in recently deglaciated areas[22,23]. The relatively consistent MIS 11 KMB silt flux may suggest persistent valley glaciers or local ice caps, or both, on that terrane, perhaps due to regionally high snow accumulation[24]. The NMB and AB silt increase at the end of MIS 11 implies south GIS regrowth, and an associated increase in subglacial erosion[8] and

©2014 Macmillan Publishers Limited. All rights reserved

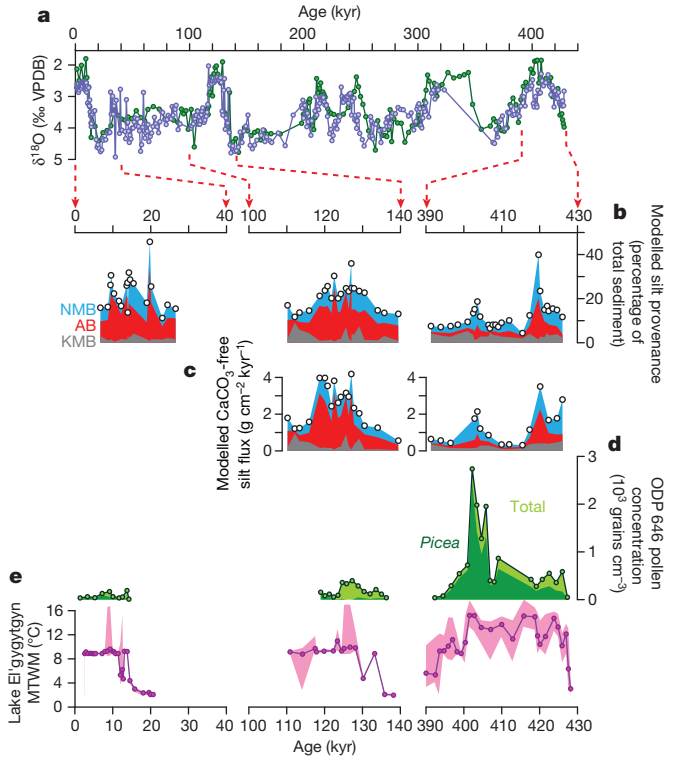


Figure 3 | Terrigenous silt provenance estimates for MD99-2227 compared with other Eirik Drift indicators of GIS extent during the Holocene, LIG and MIS 11. Age model uncertainties preclude direct assessment of millennial-scale leads and lags for the LIG and MIS 11. **a**, *Neogloboquadrina pachyderma* (s) $\delta^{18}O$ from ODP Site 646 (ref. 4; green) and MD99-2227 (Methods; blue). **b, c**, MD99-2227 median inferred silt provenance estimated from isotope mixing model, expressed as percentage of total sediment (**b**) and flux of $CaCO_3$-free silt (**c**). Holocene and LIG values are recast from Sr–Nd–Pb isotope composition in ref. 16. Holocene flux data are not presented owing to stretching in the upper 15 m of MD99-2227[19]. **d**, ODP Site 646 pollen data[4], showing *Picea* (dark green) and total (minus *Pinus*; light green) pollen concentrations. For general comparison only, we plot ODP Site 646 data on the MD99-2227 age model, using magnetic susceptibility to correlate the two cores (Extended Data Fig. 4; Methods). **e**, Reconstructed mean temperature of the warmest month (MTWM) at Lake El'gygytgyn, Russia[5]. Upper and lower bounds of shaded interval denote maximum and minimum warmest-month reconstructed temperatures, respectively.

glaciofluvial sediment transport[18,22,23] on those terranes. Ablation would subsequently decrease as climate further cooled into the next glacial stage, leading to reduced subglacial erosion and meltwater discharge. This pattern is consistent with lower Greenland Precambrian silt sedimentation at the end of MIS 11, and is similar to trends at the end of the LIG[16] (Fig. 3b, c).

An Eirik Drift pollen record from Ocean Drilling Program (ODP) Site 646 (Fig. 1) shows the highest pollen concentrations of the past ~1,000 kyr and abundant *Picea* (spruce) pollen during the $\delta^{18}O$ minimum of MIS 11[4] (Fig. 3d), but these data do not relate directly to GIS extent and volume. Sedimentary ancient DNA recovered from silty ice at the base of the Dye-3 ice core (Fig. 1) is evidence for ice-free, forested conditions at one location on south Greenland[3], but age estimates for the DNA-bearing silty ice span ~400 kyr to >1,000 kyr and thus preclude direct assignment to MIS 11. We now show that south Greenland was nearly completely deglaciated during MIS 11, providing the ice-free conditions required for development of boreal forest[3,4]. In contrast, basal ice from Summit Station (Fig. 1) probably pre-dates ~400 kyr ago[3,25], indicating survival of the central GIS during MIS 11. Similarly, continuous MIS 11 deposition of ice-rafted debris in the western Iceland Sea, following the Termination V peak in ice-rafted debris across the polar North Atlantic, suggests the presence of marine-terminating ice on east-central and northeast Greenland during MIS 11[26,27].

We compare these constraints on MIS 11 GIS extent (see, for example, Fig. 1b) with ice-sheet models that simulate GIS response to past and future climate warming, to qualitatively assess potential GIS contribution to MIS 11 sea level. Models with ice absence on the NMB and AB terranes simulate a GIS contribution to sea-level rise of >4.5 m (refs 9–12), whereas the maximum modelled GIS retreat consistent with our interpretation of an ice-free AB and NMB and other geological constraints[3,25–27] yields ~6 m of GIS sea-level contribution[7]. South GIS retreat could therefore account for all, or most, of the minimum estimated MIS 11 sea-level highstand of ~6 m (ref. 1) and is within the 1σ uncertainty envelope of the Red Sea sea-level record[2], implying a relatively stable AIS. Alternatively, the upper end of MIS 11 sea-level estimates implies complete collapse of the West AIS and at least some mass loss from the East AIS[12], in addition to 4.5–6 m of GIS sea-level contribution.

Evidence for deglaciation of south Greenland consequently suggests that a climate/ice-stability threshold was crossed during MIS 11. Coupled climate/ice-sheet models simulate a threshold for complete GIS melting of 1.7–2.0 °C of warming relative to the preindustrial late Holocene if positive dynamic feedbacks are considered[6,7]. They also simulate a long GIS response time for temperature increases approaching the 1.7–2.0 °C threshold. However, the GIS response is considerably more rapid under ~4 °C warming scenarios, with complete collapse in ~8 kyr. Unfortunately, available proxy records offer conflicting evidence for the Arctic palaeoclimate during MIS 11. The few existing temperature reconstructions from near central-south Greenland suggest that MIS 11 climate was no warmer than peak early Holocene or LIG climate[4,26]. In contrast, Siberian pollen records[5] suggest 4–5 °C of summer warming relative to the early Holocene and LIG (Fig. 3e). Limited field evidence indicates that continuous permafrost may have thawed during MIS 11 at a site in Siberia[28], although permafrost persisted locally through this interval in northern Canada[29]. In the Arctic Ocean, faunal records[30] imply that substantially warmer-than-present summer sea surface temperatures and seasonal—not perennial—sea-ice cover characterized MIS 11.

Our results therefore prompt us to ask why south Greenland deglaciated almost completely during MIS 11, when ice persisted through the LIG and the Holocene[16]. Anomalous MIS 11 Arctic warmth[5] and a seasonally ice-free ocean[30] could have pushed the GIS past an ice-stability threshold[6,7] that was not crossed during the LIG or early Holocene. Alternatively, protracted warmth during the long MIS 11 interglaciation may have allowed the GIS to respond fully to interglacial conditions that were similar to the early Holocene and LIG, with these later interglaciations being too short for comparable ice-sheet collapse. A more extensive network of Pleistocene climate and ice-extent records around Greenland is required to assess these alternative scenarios, which have important implications for predicting the long-term behaviour of the GIS in response to future climate change scenarios[6,7]. In this context, our evidence for a late-Quaternary collapse of the south GIS provides an important example of geologically recent ice-sheet instability and retreat under climate conditions within the range of those anticipated by the end of this century.

METHODS SUMMARY

We supplemented an earlier compilation[16] of Greenland stream sediment isotope geochemistry with an additional 31 samples of fine-grained glaciofluvial sediment collected from southwest Greenland (Extended Data Fig. 3). Chemical separation and isotopic analyses were conducted at the University of Wisconsin-Madison Radiogenic Isotope Laboratory. Subsamples of the 3–63 μm fraction of MIS 11 sediment from core MD99-2227 were twice leached with 0.1 M HCl to remove biogenic carbonate and authigenic Fe–Mn coatings. All samples were spiked with mixed Rb–Sr and rare-earth-element tracers, and then digested in Parr pressure vessels with hydrofluoric acid and HNO_3. Pb, Rb–Sr and rare-earth-element separation was by sequential ion-exchange column chromatography. Isotope ratios were measured on a VG Sector 54 thermal ionization mass spectrometer, with Rb, Sr, Nd and

©2014 Macmillan Publishers Limited. All rights reserved

Sm concentrations determined by isotope dilution. Procedural blanks were negligible compared with sample size.

The isotope mixing model has four terrane components (KMB, AB, NMB and Palaeogene volcanics) and five endmembers ($^{87}Sr/^{86}Sr$, ε_{Nd}, $^{206}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$). The model was run 10,000 times, with endmember composition and concentration randomly selected within the 16.5–83.5% quantile range of the stream sediment data sets for each terrane. Mixing model results are corrected for $CaCO_3$ content and presented as the median of all valid solutions returned by the 10,000 model runs, with 67% and 95% confidence intervals based on the 83.5–16.5% and 97.5–2.5% quantiles of all valid solutions, respectively (Extended Data Fig. 2).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 November 2013; accepted 2 May 2014.

1. Raymo, M. E. & Mitrovica, J. X. Collapse of polar ice sheets during the stage 11 interglacial. *Nature* **483,** 453–456 (2012).
2. Rohling, E. J. *et al.* Comparison between Holocene and Marine Isotope Stage-11 sea-level histories. *Earth Planet. Sci. Lett.* **291,** 97–105 (2010).
3. Willerslev, E. *et al.* Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317,** 111–114 (2007).
4. de Vernal, A. & Hillaire-Marcel, C. Natural variability of Greenland climate, vegetation, and ice volume during the past million years. *Science* **320,** 1622–1625 (2008).
5. Melles, M. *et al.* 2.8 Million years of Arctic climate change from Lake El'gygytgyn, NE Russia. *Science* **337,** 315–320 (2012).
6. Ridley, J., Gregory, J. M., Huybrechts, P. & Lowe, J. Thresholds for irreversible decline of the Greenland ice sheet. *Clim. Dyn.* **35,** 1049–1057 (2010).
7. Robinson, A., Calov, R. & Ganopolski, A. Multistability and critical thresholds of the Greenland ice sheet. *Nature Clim. Change* **2,** 429–432 (2012).
8. Cowton, T., Nienow, P., Bartholomew, I., Sole, A. & Mair, D. Rapid erosion beneath the Greenland ice sheet. *Geology* **40,** 343–346 (2012).
9. Cuffey, K. M. & Marshall, S. J. Substantial contribution to sea-level rise during the last interglacial from the Greenland Ice Sheet. *Nature* **404,** 591–594 (2000).
10. Huybrechts, P. Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quat. Sci. Rev.* **21,** 203–231 (2002).
11. Tarasov, L. & Peltier, W. R. Greenland glacial history, borehole constraints, and Eemian extent. *J. Geophys. Res.* **108,** 2143 (2003).
12. Lhomme, N., Clarke, G. K. C. & Marshall, S. J. Tracer transport in the Greenland Ice Sheet - constraints on ice cores and glacial history. *Quat. Sci. Rev.* **24,** 173–194 (2005).
13. Fretwell, P. Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *Cryosphere* **7,** 375–393 (2013).
14. Bamber, J. L. *et al.* A new bed elevation dataset for Greenland. *Cryosphere* **7,** 499–510 (2013).
15. Fagel, N. & Hillaire-Marcel, C. Glacial/interglacial instabilities of the Western Boundary Under Current during the last 365 kyr from Sm/Nd ratios of the sedimentary clay-size fractions at ODP site 646 (Labrador Sea). *Mar. Geol.* **232,** 87–99 (2006).
16. Colville, E. J. *et al.* Sr-Nd-Pb isotope evidence for ice-sheet presence on southern Greenland during the last interglacial. *Science* **333,** 620–623 (2011).
17. Stoner, J. S., Channell, J. E. T. & Hillaire-Marcel, C. Magnetic properties of deep-sea sediments off southwest Greenland: evidence for major differences between the last two deglaciations. *Geology* **23,** 241–244 (1995).
18. Carlson, A. E., Stoner, J. S., Donnelly, J. P. & Hillaire-Marcel, C. Response of the southern Greenland Ice Sheet during the last two deglaciations. *Geology* **36,** 359–362 (2008).
19. Evans, H. F. *et al.* Paleointensity-assisted chronostratigraphy of detrital layers on the Eirik Drift (North Atlantic) since marine isotope stage 11. *Geochem. Geophys. Geosyst.* **8,** Q11007 (2007).
20. Hillaire-Marcel, C., de Vernal, A. & McKay, J. Foraminifer isotope study of the Pleistocene Labrador Sea, northwest North Atlantic (IODP Sites 1302/03 and 1305), with emphasis on paleoceanographical differences between its ''inner'' and ''outer'' basins. *Mar. Geol.* **279,** 188–198 (2011).
21. Hillaire-Marcel, C., de Vernal, A., Bilodeau, G. & Wu, G. Isotope stratigraphy, sedimentation rates, deep circulation, and carbonate events in the Labrador Sea during the last ~200 ka. *Can. J. Earth Sci.* **31,** 63–89 (1994).
22. Hasholt, B. in *Proc. Int. Symp. Erosion Sediment Yield: Global Regional Perspectives* (eds Walling, D. E. & Webb, B. W.) 105–114 (IAHS-AISH Publication, 1996).
23. Church, M. & Ryder, J. M. Paraglacial sedimentation: a consideration of fluvial processes conditioned by glaciation. *Geol. Soc. Am. Bull.* **83,** 3059–3071 (1972).
24. Ohmura, A. & Reeh, N. New precipitation and accumulation maps for Greenland. *J. Glaciol.* **37,** 140–148 (1991).
25. Bierman, P. R. *et al.* Preservation of a preglacial landscape under the center of the Greenland ice sheet. *Science* **344,** 402–405 (2014).
26. Helmke, J. P. & Bauch, H. A. Comparison of conditions between the polar and subpolar North Atlantic region over the last five climate cycles. *Paleoceanography* **18,** 1036 (2003).
27. Bauch, H. A. Interglacial climates and the Atlantic meridional overturning circulation: is there an Arctic controversy? *Quat. Sci. Rev.* **63,** 1–22 (2013).
28. Vaks, A. *et al.* Speleothems reveal 500,000-year history of Siberian permafrost. *Science* **340,** 183–186 (2013).
29. Froese, D. G., Westgate, J. A., Reyes, A. V., Enkin, R. J. & Preece, S. J. Ancient permafrost and a future, warmer Arctic. *Science* **321,** 1648 (2008).
30. Cronin, T. M. *et al.* A 600-ka Arctic sea-ice record from Mendeleev Ridge based on ostracodes. *Quat. Sci. Rev.* **79,** 157–167 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. de Vernal for access to pollen records and archived sediment; C. Hillaire-Marcel for discussions of MD99-2227 stratigraphy and geochemistry; J. Briner, B. Hudson, S. Kelley and N. Larsen for providing samples; and E. Colville, P. Holm and S. Strano for assistance in the field. This research was supported by US NSF awards ANS-0902571 (A.E.C., B.L.B.) and -0902751 (J.S.S.), and a Canadian NSERC fellowship (A.V.R.).

Author Contributions A.E.C., B.L.B. and J.S.S. had the idea for the study; A.V.R., A.E.C. and R.G.H. designed and conducted field research in Greenland; B.W. conducted grain-size analysis; A.V.R. and B.L.B. conducted isotopic analyses; K.W. sampled and identified foraminifera; A.V.R. and D.J.U. implemented the isotope mixing model; A.E.C., R.G.H., J.S.S. and K.W. developed the age model for MD99-2227; A.V.R., A.E.C. and B.L.B. synthesized the results; and A.V.R. and A.E.C. wrote the manuscript, with input from all authors.

Author Information Greenland stream sediment and MD99-2227 data have been deposited with the NOAA National Climatic Data Center (http://hurricane.ncdc.noaa.gov/pls/paleox/f?p=519:1:0::::P1_STUDY_ID:16436). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.V.R. (areyes@ualberta.ca) or A.E.C. (acarlson@coas.oregonstate.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13282

Purkinje-cell plasticity and cerebellar motor learning are graded by complex-spike duration

Yan Yang[1] & Stephen G. Lisberger[1,2]

Behavioural learning is mediated by cellular plasticity, such as changes in the strength of synapses at specific sites in neural circuits. The theory of cerebellar motor learning[1–3] relies on movement errors signalled by climbing-fibre inputs to cause long-term depression of synapses from parallel fibres to Purkinje cells[4,5]. However, a recent review[6] has called into question the widely held view that the climbing-fibre input is an 'all-or-none' event. In anaesthetized animals, there is wide variation in the duration of the complex spike (CS) caused in Purkinje cells by a climbing-fibre input[7]. Furthermore, the amount of plasticity in Purkinje cells is graded according to the duration of electrically controlled bursts in climbing fibres[8,9]. The duration of bursts depends on the 'state' of the inferior olive and therefore may be correlated across climbing fibres[8,10]. Here we provide a potential functional context for these mechanisms during motor learning in behaving monkeys. The magnitudes of both plasticity and motor learning depend on the duration of the CS responses. Furthermore, the duration of CS responses seems to be a meaningful signal that is correlated across the Purkinje-cell population during motor learning. We suggest that during learning, longer bursts in climbing fibres lead to longer-duration CS responses in Purkinje cells, more calcium entry into Purkinje cells, larger synaptic depression, and stronger learning. The same graded impact of instructive signals for plasticity and learning might occur throughout the nervous system.

We recorded the neural activity of single Purkinje cells in the floccular complex of the cerebellum in monkeys that had been trained to perform smooth-pursuit eye movements[11,12]. We chose to study the floccular complex because its output drives pursuit eye movements[13] as well as pursuit learning[11,14] through a disynaptic pathway to extraocular motoneurons[15].

During pursuit eye movements, floccular Purkinje cells show direction tuning in simple-spike firing, which is driven by mossy-fibre inputs to the cerebellum. To set the stage for learning experiments, we defined each neuron's 'on direction' by the pursuit direction with the largest increases in simple-spike firing rate; pursuit in the opposite, or 'off direction', was associated with decreases in simple-spike firing rate (Extended Data Fig. 1b, c). The CS responses driven by climbing-fibre inputs show direction tuning opposite to that for the simple-spike responses (Extended Data Fig. 1).

Motor learning occurred during a succession of 'trials'. In each trial, monkeys tracked a target that first moved for 250 ms at 20 deg per s in a 'pursuit' direction orthogonal to the on-direction of the Purkinje cell under study (horizontal in Fig. 1b). Then, the target underwent an instructive change in direction that added orthogonal target motion at 30 deg per s for 400 ms. The 'instruction' was in either the on or off direction for the simple-spike response of the Purkinje cell under study (downward or upward in Fig. 1b). After the instruction ended, the target continued to move in the pursuit direction for 200 ms and then stopped.

We used a 'random-direction learning paradigm'[12] (Fig. 1a) to study neural correlates of motor learning on the time scale of single behavioural trials. Each eye-velocity response showed a small, transient, learned deflection in the direction of the instruction on the prior trial (Fig. 1b, arrowhead on vertical-eye-velocity traces). The small upward deflection of vertical eye velocity in the nth trial (Fig. 1b, blue traces) was caused by the upward instruction on the prior, $n-1$ trial (red traces). The small downward deflection at the same time on the $n-1$ trial (red traces) was caused by the downward instruction on the $n-2$ trial. We call these brief deflections 'trial-over-trial' behavioural learning.

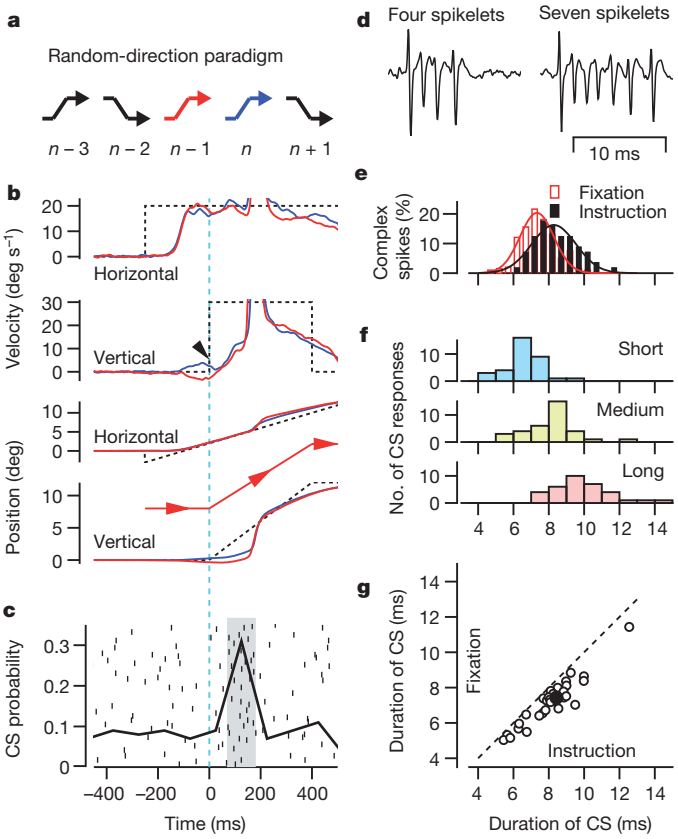


Figure 1 | Design of the learning paradigm and variation of CS duration. **a**, Trajectories of target position during a random-order learning block. **b**, Trial-over-trial learning in an example pair of consecutive trials. From top to bottom, superimposed traces are horizontal and vertical eye velocity, and horizontal and vertical eye position. Red and blue traces show data for trials '$n-1$' and 'n'. Colours of arrows in **a** refer to colours of traces in **b.** Dashed traces show target motion. The red zigzag between the position traces shows the trajectory of target position in two dimensions. The arrowhead on the vertical-eye-velocity trace indicates the learned response, and the light blue dashed line shows the time of the instruction. **c**, Times of occurrence of CS responses and the probability of CS responses in 100-ms bins. The grey shaded area shows the analysis window for CS responses to the instruction. **d**, Two example CS waveforms from an example Purkinje cell. **e**, Superimposed, statistically different distributions of CS duration during fixation versus learning for the example Purkinje cell (two-tailed t-test, $P < 0.01$). **f**, Distributions of mean CS duration for the short-, medium- and long-duration CS responses across the population of 34 Purkinje cells. **g**, Scatter plot shows average CS duration during fixation versus learning for all 34 Purkinje cells. The filled symbol indicates the example neuron in **e**. The dashed black line has a slope of one.

[1]Department of Neurobiology, Duke University, Durham, North Carolina 27710, USA. [2]Howard Hughes Medical Institute, Duke University, Durham, North Carolina 27710, USA.

©2014 Macmillan Publishers Limited. All rights reserved

For each of 34 individual Purkinje cells with well-isolated CS waveforms, we observed CS responses with a probability of 0.3 to 0.4 in the interval from 75 to 175 ms after an instruction (Fig. 1c), but a wide range in the total duration of the extracellular CS waveform (Fig. 1d) and in the number of spikelets[7]. We found a linear relationship between CS duration and the number of spikelets (Extended Data Fig. 2), and a strong correlation. We assume that the duration of the extracellular CS waveform is a valid probe for the duration of the depolarization in the Purkinje cell's membrane potential[8]. Our results show an effect of CS duration on plasticity and therefore support this assumption, with the caveat that it is difficult to be sure of intracellular events from extracellular recordings.

The duration of CS responses varied widely even in a given behavioural condition and differed between behavioural conditions, that is, fixation versus learning (Fig. 1e, g). To study the effects of CS duration, we divided the distribution for each neuron into thirds by trisecting it at the mean ± 0.44 s.d.; we categorized the duration of the CS in instruction trials as 'short', 'medium' or 'long'. The mean CS durations in the three groups were 6.59 ms, 8.12 ms and 9.84 ms across our sample of Purkinje cells (Fig. 1f).

Our data analysis monitors plasticity linked to the CS responses driven by natural sensory stimuli during behavioural learning. We broke the string of learning trials into pairs of 'instruction' and 'test' trials (Fig. 2a). We included only pairs with an instruction in the off direction for simple-spike firing in the first trial; instructions in the on direction seldom caused a CS response (Extended Data Fig. 1b). To demonstrate the trial-over-trial effects linked to the duration of a CS response, we grouped the pairs of trials according to whether the duration of the CS in the instruction trial was short, medium, long or absent. Within each group, we then averaged the trial-over-trial change in firing rate between the instruction and test trial.

Simple-spike firing rate underwent trial-over-trial depression that was related to the presence[11,16] and duration of the CS response in the instruction trial (Fig. 2b). The trial-over-trial depression was largest after a long-duration CS (Fig. 2b, red trace), smallest after a short-duration CS (blue trace), and intermediate after a medium-duration CS (Fig. 2b, yellow trace). Trial-over-trial changes of simple-spike firing rate showed a small potentiation that was not significantly different from zero (two-sided Wilcoxon signed-rank test, $P > 0.05$) if the off-direction instruction failed to evoke a CS response (Fig. 2b, black trace)[11,16].

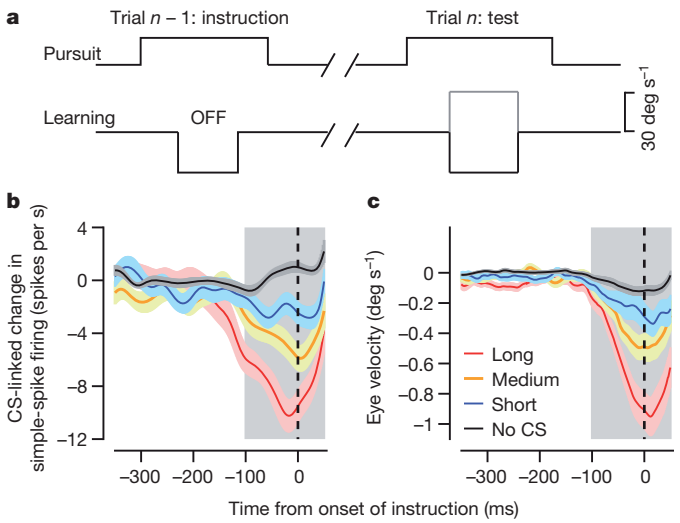


Figure 2 | Effect of CS duration on trial-over-trial depression and learning. **a**, The schematic shows the sequence of instruction and test trials. The pursuit stimulus was the same in both trials, and the instruction provided an off-direction target motion. **b**, **c**, Plots show trial-over-trial changes in simple-spike firing rate (**b**) or eye velocity (**c**). Red, yellow, blue and black traces show data when the CS response to the instruction had a long, medium or short duration, or was absent. Shaded areas around the traces show standard errors of the mean of firing rate averaged across Purkinje cells. The vertical dashed line shows the time of the instruction; grey shading shows the analysis interval used for Fig. 3. $N = 34$ Purkinje cells.

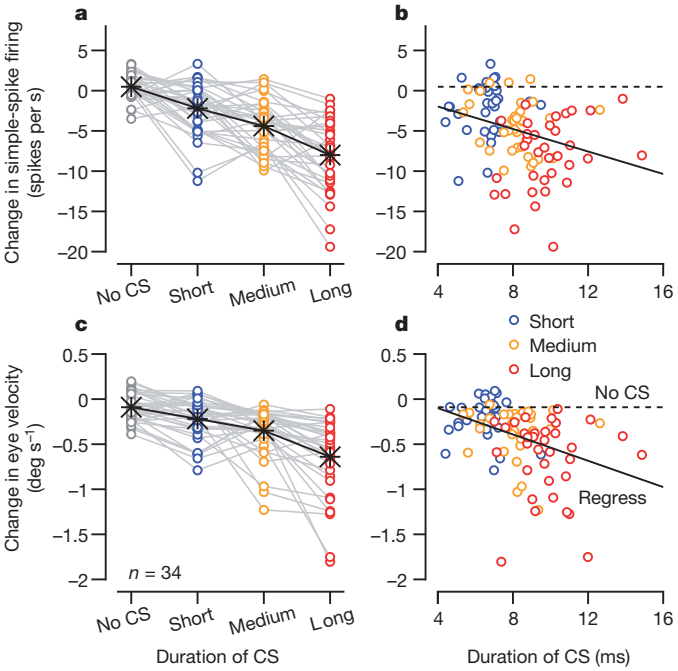


Figure 3 | Quantitative analysis of the effects of CS duration for all individual Purkinje cells. **a**, **b**, CS-linked trial-over-trial change in simple-spike firing rate. **c**, **d**, CS-linked trial-over-trial learning in eye velocity. Different colours indicate the absence or duration of the CS responses to an off-direction instruction. In **a** and **c**, grey lines connect four symbols for each neuron; black symbols show means across the population. In **b** and **d**, CS-linked trial-over-trial changes are plotted as a function of the mean duration of the CS waveforms for each group. The continuous lines were obtained by linear regression. The horizontal dashed lines show the averages for the No-CS group. $N = 34$ Purkinje cells.

CS duration also affected the trial-over-trial learning in eye velocity (Fig. 2c). Learning was strongest if the CS duration was long, became progressively smaller when CS duration was medium or short, and was very weak in the absence of a CS response to the off-direction instruction. In contrast, the duration of the post-CS pause in simple-spike firing rate did not affect the magnitudes of trial-over-trial depression in simple-spike firing rate or trial-over-trial learning in eye velocity (Extended Data Fig. 3).

The data from all 34 Purkinje cells show a consistent trend of downward slopes as a function of whether the CS response is short, medium or long (Fig. 3a, c). For statistical analysis, we plotted the magnitude of the trial-over-trial effects in each neuron as a function of the mean of CS duration in that neuron's short, medium and long groups (Fig. 3b, d). Regression analysis yielded negative slopes. The slope for firing rate was −0.70 spikes per s for each millisecond of CS duration (95% confidence limits on slope −1.13 to −0.27). The slope for eye velocity was −0.07 deg per s for each ms of CS duration (95% confidence limits on slope −0.11 to −0.04).

We next evaluated the mechanisms for the effect of CS duration on trial-over-trial depression of simple-spike firing rate. We averaged absolute simple-spike firing rate as a function of time for different groups of trials (Fig. 4). The traces all start with an increase in firing rate of 10–15 spikes per s at about −150 ms (Fig. 4a, arrow). This increase is related to the onset of eye velocity in the pursuit direction, which is orthogonal to the on direction; the response to pursuit in the on direction would have been much larger. The key interval for probing CS-linked trial-over-trial simple-spike depression is later, in the grey shaded area from 100 ms before to 50 ms after the onset of the instruction. CS responses would have occurred still later, after the end of the traces.

©2014 Macmillan Publishers Limited. All rights reserved

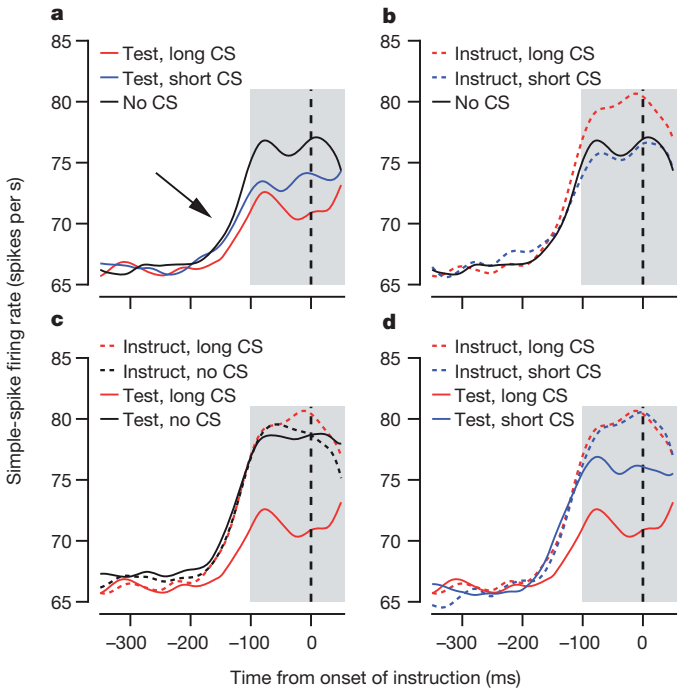


Figure 4 | Evidence that CS duration affects the magnitude of CS-linked plasticity. All graphs plot simple-spike firing rate as a function of time from the onset of the instruction. **a**, Test trials that followed instruction trials with CS responses of different durations. The diagonal arrow points out the increase in firing rate associated with the initiation of pursuit in the pursuit direction. **b**, Instruction trials with long, short or absent CS responses. **c**, Pairs of trials that have the same simple-spike firing rates in the instruction trials but differ in the presence versus absence of a CS response to the instruction. **d**, Pairs of trials that have the same simple-spike firing rates in the instruction trials but differ in the duration of the CS response to the instruction. The grey shaded areas show the intervals during which it is valid to evaluate the effects of plasticity. $N = 34$ Purkinje cells.

Evidence for plasticity of simple-spike firing comes from the observation that a long-duration CS on the instruction trials drove the simple-spike firing rate on the subsequent test trials below the baseline established by trials without a CS (Fig. 4a, red versus black traces). Short-duration CS responses also drove simple-spike firing rate below the baseline (Fig. 4a, blue versus black traces). Thus, trial-over-trial depression must represent an active change, and cannot be due simply to the loss of a slight enhancement of simple-spike firing rate at the time of a CS response[16].

The elevated simple-spike firing rate on instruction trials with a CS of long duration (Fig. 4b, red dashed trace versus blue or black) raises the possibility that the higher simple-spike firing rate on the instruction trial could be the causal agent that drives trial-over-trial depression of simple-spike firing rate[17]. We rejected this possibility by comparing two groups of instruction–test pairs that had equal simple-spike firing rates, but differed in the presence, duration or absence of a CS in response to the instruction (Fig. 4c, d). Simple-spike firing rate showed trial-over-trial depression only if there was a CS on the instruction trial (Fig. 4c), and the magnitude of the depression depended on the duration of the CS (Fig. 4d). Simple-spike firing rate on the instruction trial controls neither CS duration nor plasticity.

It is the context of the effects we demonstrate that is most noteworthy. We have been able to monitor cellular events that actually occur during learning. Our approach analyses the effect of natural variation in CS duration on trial-over-trial changes in simple-spike firing rate and behavioural performance. Of course, longer-duration CS responses should allow more calcium to enter a Purkinje cell and promote greater plasticity[18], as demonstrated by electrical stimulation in the inferior olive[9]. In the present study, we take the next step by showing that CS duration matters for plasticity of simple-spike firing rate during motor learning driven by natural sensory stimuli.

The effect of CS duration on trial-over-trial learning in eye velocity has important implications for the organization of motor learning in the cerebellum. Learned eye velocity is a measure of the combined plasticity of simple-spike firing rate of all Purkinje cells in the floccular complex. If CS duration and the magnitude of plasticity varied randomly across the population of Purkinje cells on a given trial, then the trial-over-trial change in eye velocity would not depend on CS duration. As learning is much greater when the Purkinje cell under study has a long CS, CS duration seems to be coordinated across the population of Purkinje cells.

Coordinated modulation of CS duration across the population of Purkinje cells could control the amount of learning induced by a given motor error. Descending influences on the inferior olive or local Purkinje-cell-membrane excitability could mediate the modulation. We lean towards modulation at the level of the inferior olive. The number of spikes in a climbing-fibre burst reads out the state of the inferior olive[8,10], and is one determinant of CS duration[8]. Thus, intentional modulation of the state of the inferior olive could correlate the durations of bursts across a large group of climbing fibres. Modulation could arise either from the cerebral cortex, or from the cerebellum through neurons in the deep cerebellar nuclei that inhibit the inferior olive[19].

Local modulation in the cerebellar cortex could affect CS duration by controlling the excitability of the Purkinje-cell membrane[20,21]. If modulation occurs locally, then correlations in CS duration across the population could arise through neuron–neuron correlations in simple-spike firing rate[22], or through coordinated sensory inputs[23]. We do not know whether simple-spike firing rate is a good index of Purkinje-cell membrane excitability *in vivo*. If it is, then our data argue against local control of CS duration because CS duration grades simple-spike plasticity even when instruction trials are selected so that simple-spike firing rate is equated (Fig. 4d).

Intentional modulation of CS duration could be important in cerebellar function. Purkinje cell output may modulate the duration of its own CS inputs and restrict its own learning range[16,24–26], limiting short-term learning in the cerebellar cortex, and enabling transfer of plasticity to the deep cerebellar nucleus[27]. Modulation of the amount of learning induced by a given instructive stimulus might facilitate cerebellar learning under conditions that are particularly important to the organism ('Coach' is screaming at you during practice), and prevent learning when it would be counter-productive (in the middle of a crucial play in a game). Indeed, we observed that learning occurs more quickly and the durations of CS responses are generally longer in the first 30 trials of a learning block when the same instruction occurs in 100 consecutive trials. Such a situation was suggested before on theoretical grounds[28].

The question of what happens in our brains while we learn is answered in the most detail for cerebellar motor learning. But the principles of cerebellar learning almost certainly will apply to other learning systems. Perhaps long-term potentiation in the hippocampus can be graded by modulation from the cerebral cortex. In the cerebellum, the duration of the climbing fibre input is particularly accessible to top-down control, but the same effect could occur in other systems through inhibition, for example. Thus, the details of motor learning in the cerebellum may lead to principles that apply in all learning systems in the brain.

METHODS SUMMARY

We performed experiments on two adult male rhesus monkeys using techniques that have been described in detail previously[11,12,16]. The data in this paper are drawn from a single data set that also provides the data for two other papers that cover different aspects of the results[16,29]. All procedures followed a protocol that had been approved in advance by the Institutional Animal Care and Use Committee at the University of California San Francisco (UCSF), where the experiments were performed. All procedures were in accordance with the US National Institutes of Health (NIH) Guild for the Care and Use of Laboratory Animals.

Monkeys were prepared for recordings of eye movements and single neurons, and were trained to fixate and track targets that appeared on a video screen in front

©2014 Macmillan Publishers Limited. All rights reserved

of them. In each daily experiment, we isolated the extracellular potentials from an individual Purkinje cell in the floccular complex of the cerebellum, and recorded from that Purkinje cell during many repetitions of a few carefully chosen target motions. Data were recorded for analysis after the experiment. We used homemade software to identify each simple spike and CS through inspection of the records on a video screen. Measurements of CS duration were made with the experimenter blind to the context of the CS and to the eye velocity and simple-spike firing rate at the time.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 December 2013; accepted 25 March 2014.
Published online 11 May 2014.

1. Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* **202,** 437–470 (1969).
2. Albus, J. S. A theory of cerebellar function. *Math. Biosci.* **10,** 25–61 (1971).
3. Ito, M. Neural design of the cerebellar motor control system. *Brain Res.* **40,** 81–84 (1972).
4. Gilbert, P. F. & Thach, W. T. Purkinje cell activity during motor learning. *Brain Res.* **128,** 309–328 (1977).
5. Ito, M. & Kano, M. Long-lasting depression of parallel fiber–Purkinje cell transmission induced by conjunctive stimulation of parallel fibers and climbing fibers in the cerebellar cortex. *Neurosci. Lett.* **33,** 253–258 (1982).
6. Najafi, F. & Medina, J. F. Beyond "all-or-nothing" climbing fibers: graded representation of teaching signals in Purkinje cells. *Front. Neural Circ.* **7** (2013).
7. Maruta, J., Hensbroek, R. A. & Simpson, J. I. Intraburst and interburst signaling by climbing fibers. *J. Neurosci.* **27,** 11263–11270 (2007).
8. Mathy, A. *et al.* Encoding of oscillations by axonal bursts in inferior olive neurons. *Neuron* **62,** 388–399 (2009).
9. Rasmussen, A. *et al.* Number of spikes in climbing fibers determines the direction of cerebellar learning. *J. Neurosci.* **33,** 13436–13440 (2013).
10. Bazzigaluppi, P. *et al.* Olivary subthreshold oscillations and burst activity revisited. *Front. Neural Circ.* **6,** 91 (2012).
11. Medina, J. F. & Lisberger, S. G. Links from complex spikes to local plasticity and motor learning in the cerebellum of awake-behaving monkeys. *Nature Neurosci.* **11,** 1185–1192 (2008).
12. Yang, Y. & Lisberger, S. G. Learning on multiple timescales in smooth pursuit eye movements. *J. Neurophysiol.* **104,** 2850–2862 (2010).
13. Zee, D. S., Yamazaki, A., Butler, P. H. & Gücer, G. Effects of ablation of flocculus and paraflocculus of eye movements in primate. *J. Neurophysiol.* **46,** 878–899 (1981).
14. Kahlon, M. & Lisberger, S. G. Changes in the responses of Purkinje cells in the floccular complex of monkeys after motor learning in smooth pursuit eye movements. *J. Neurophysiol.* **84,** 2945–2960 (2000).
15. Highstein, S. M. Synaptic linkage in the vestibulo-ocular and cerebello-vestibular pathways to the VIth nucleus in the rabbit. *Exp. Brain Res.* **17,** 301–314 (1973).
16. Yang, Y. & Lisberger, S. G. Interaction of plasticity and circuit organization during the acquisition of cerebellum-dependent motor learning. *eLife* **2,** e01574 (2013).
17. Hartell, N. A. Strong activation of parallel fibers produces localized calcium transients and a form of LTD that spreads to distant synapses. *Neuron* **16,** 601–610 (1996).
18. Wang, S. S.-H., Denk, W. & Hausser, M. Coincidence detection in single dendritic spines mediated by calcium release. *Nature Neurosci.* **3,** 1266–1273 (2000).
19. Bengtsson, F. & Hesslow, G. Cerebellar control of the inferior olive. *Cerebellum* **5,** 7–14 (2006).
20. Callaway, J. C., Lasser-Ross, N. & Ross, W. N. IPSPs strongly inhibit climbing fiber-activated $[Ca^{2+}]_i$ increases in the dendrites of cerebellar Purkinje neurons. *J. Neurosci.* **15,** 2777–2787 (1995).
21. Kitamura, K. & Hausser, M. Dendritic calcium signaling triggered by spontaneous and sensory-evoked climbing fiber input to cerebellar Purkinje cells *in vivo*. *J. Neurosci.* **31,** 10847–10858 (2011).
22. Medina, J. F. & Lisberger, S. G. Variation, signal, and noise in cerebellar sensory-motor processing for smooth-pursuit eye movements. *J. Neurosci.* **27,** 6832–6842 (2007).
23. Najafi, F., Giavannucci, A., Wang, S. S.-H. & Medina, J. F. Sensory driven enhancement of calcium signals in individual Purkinje cell dendrites of awake mice. *Cell Rep.* **6,** 792–798 (2014).
24. Mauk, M. D. & Donegan, N. H. A model of Pavlovian eyelid conditioning based on the synaptic organization of the cerebellum. *Learn. Mem.* **4,** 130–158 (1997).
25. Kenyon, G. T., Medina, J. F. & Mauk, M. D. A mathematical model of the cerebellar-olivary system I: self-regulating equilibrium of climbing fiber activity. *J. Comput. Neurosci.* **5,** 17–33 (1998).
26. Miall, R. C., Keating, J. G., Malkmus, M. & Thach, W. T. Simple spike activity predicts occurrence of complex spikes in cerebellar Purkinje cells. *Nature Neurosci.* **1,** 13–15 (1998).
27. Ohyama, T., Nores, W. L., Medina, J. F., Riusech, F. A. & Mauk, M. D. Learning-induced plasticity in deep cerebellar nucleus. *J. Neurosci.* **26,** 12656–12663 (2006).
28. Gilbert, P. F. C. A theory of memory that explains the function and structure of the cerebellum. *Brain Res.* **70,** 1–18 (1974).
29. Yang, Y. & Lisberger, S. G. (2014) Role of plasticity at different sites across the time course of cerebellar motor learning. *J. Neurosci.* (in the press).

Acknowledgements We thank C. Hull, L. Glickfeld and R. Mooney for helpful comments, and M. Joshua and J. Lee for 'blind' tests of our measurements of complex-spike duration. Research supported by the Howard Hughes Medical Institute and by the National Eye Institute of the National Institutes of Health under Award Number R01-EY003878. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author Contributions Y.Y. performed all experiments and data analysis. Y.Y. and S.G.L. designed and interpreted experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G.L. (lisberger@neuro.duke.edu) or Y.Y. (yanyang@neuro.duke.edu).

©2014 Macmillan Publishers Limited. All rights reserved

METHODS

We performed experiments on two adult male rhesus monkeys using techniques that have been described in detail previously (refs 11, 12, 16). The data in this paper are drawn from a single data set that also provides the data for two other papers that cover different aspects of the results (refs 16 and 29). All procedures followed a protocol that had been approved in advance by the Institutional Animal Care and Use Committee at the University of California San Francisco (UCSF), where the experiments were performed. All procedures were in accordance with the US National Institutes of Health (NIH) Guild for the Care and Use of Laboratory Animals.

We used sterile procedures under surgical levels of gas anaesthesia to perform surgery from which the animals recovered, to prepare each monkey for experiments. After each surgery, monkeys received several days of treatment with systemic analgesics. In the first two surgeries, we implanted a socket on the skull to allow restraint of the monkey's head, and an eye coil on one eye to allow precise monitoring of eye position. Monkeys then were trained to fixate and track targets that appeared on a video screen in front of them. They received fluid rewards for accurate tracking, which involved keeping eye position within a small invisible rectangular window that moved with the target. Monkeys would normally work for ~2,000 rewards daily. After monkeys were fully experienced in the behavioural tasks, we performed one more surgery to make a craniotomy and implant a sealable recording cylinder to allow access to the cerebellum with microelectrodes.

An experiment consisted of discreet target motions presented in what we call 'trials'. Each trial started when the monkey fixated a 0.3-deg white spot at the centre of the video screen. After the monkey had established stable fixation, the spot expanded to 0.5 deg, displaced by a few degrees to a new location and started to move at constant speed. It then underwent one or two more changes in direction and speed before stopping at an eccentric location. The details of the target motion were selected to match the preferred direction of the Purkinje cell under study, as well as the exact goals of each experiment (see full text of paper and Extended Data Fig. 1).

In each daily experiment, we introduced a microelectrode into the floccular complex of the cerebellum, isolated the extracellular potentials from individual Purkinje cells, and recorded from that Purkinje cell during many repetitions of a few carefully chosen target motions. We took special care to obtain excellent isolation of the CS and simple-spike waveforms. Data were recorded for analysis after the experiment. After the experiment, we displayed the data from each behavioural trial on a video screen and used custom-made software to identify each simple spike and CS. We obtained records of simple-spike firing rate for individual trials with an inverse interspike-interval algorithm followed by smoothing with a Gaussian function ($\delta = 15$ ms).

Four techniques require extra explanation. First, we needed to select Purkinje cell recordings for analysis on the basis of how well the CS responses were isolated. It is a challenge to maintain excellent isolation of CS responses through ~400 trials of a random-direction learning block. As a result, we were able to identify the CS responses in 118 Purkinje cells, but we were able to measure the duration of the waveforms throughout the recordings in only 34 Purkinje cells. The selection of Purkinje cells was performed on the basis of the CS waveforms alone, before we had attempted any analysis of an effect of CS duration. The subjective procedure we used to select the Purkinje cells for analysis was analogous to the subjective procedures used in all extracellular single-unit recordings from awake behaving animals. Before including data in an analysis, it is always necessary to decide whether or not the isolation of unitary spikes is good enough to allow meaningful conclusions.

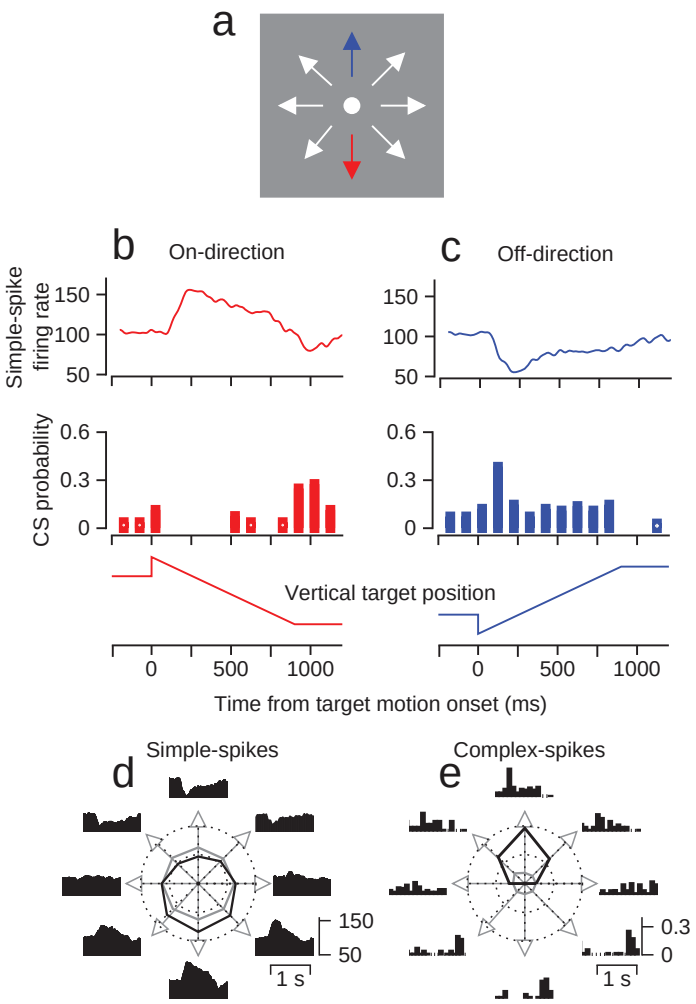
Second, the veracity of our observations depends on the objectivity and accuracy of our measurements of CS duration. The measurements were made manually by one of the authors. As detailed in Extended Data Fig. 2, CS duration was measured from the beginning of the first deflection of the extracellular potential to the time of the return to baseline potential (Extended Data Fig. 2, downward arrows on each trace). The investigator who analysed the data was blind to whether the CS under measurement occurred during fixation or in response to an instruction. Also, all measurements of CS duration were made before a given neuron's simple-spike firing rate and the associated eye velocities were analysed at all. The observer was blind to simple-spike firing rate and eye velocity in the current trial or the next trial. Thus, there was no chance of unconscious bias. To be safe, we also asked two laboratory members to measure CS duration for several representative Purkinje cells. Their measurements of CS duration agreed well statistically with those of the primary observer. We also note that random errors in measuring CS duration should dilute the effects we report here, and that shuffling the CS durations eliminated any effects of CS duration on trial-over-trial changes in simple-spike firing rate or trial-over-trial learning in eye velocity.

Our main data-analysis approach was to evaluate the trial-over-trial changes in either simple-spike firing rate or eye velocity between pairs of trials. In each pair, we considered the first trial to be the 'instruction' trial and the second trial to be the 'test' trial. The time between the two trials was ~2.5 s. We selected only the pairs of trials with instructions on the first trial that were in the off direction for simple-spike firing of the Purkinje cell under study. We then divided the pairs into groups according to the presence, duration or absence of a CS response on the instruction trial. For each pair, we calculated the difference between the simple-spike firing rate (eye velocity) on the instruction trial and on the test trial for each millisecond from 350 ms before to 50 ms after the time of the instruction. We then averaged across the trials in each group of pairs for each Purkinje cell to obtain estimates of the trial-over-trial change in the time varying simple-spike firing rate or eye velocity. For many of our analyses, we then averaged across 34 Purkinje cells as well.

Finally, for the analysis in Fig. 4c, d, we needed to select pairs of trials. We did so to achieve the specific criterion of matching the simple-spike firing rates in groups of instruction trials that differed according to the duration or presence versus absence of a CS response. To create the required match, we analysed each Purkinje cell separately and used the full sample of pairs of trials with long CS duration. From the pairs of trials with no CS or short CS duration in the instruction trial in each neuron, we selected a subset according to the simple-spike firing rate in the instruction trial. We included pairs of trials from that group only if simple-spike firing rate in the instruction trial was within the range defined by the mean plus or minus one standard deviation of the simple-spike firing rate in instruction trials with a long CS duration. We then created averages of simple-spike firing rate in the instruction and test trials that met our selection criteria for each neuron, and averaged the results across Purkinje cells. This enabled us to assess whether CS presence and duration had effects that were independent of simple-spike firing rate on the instruction trial.

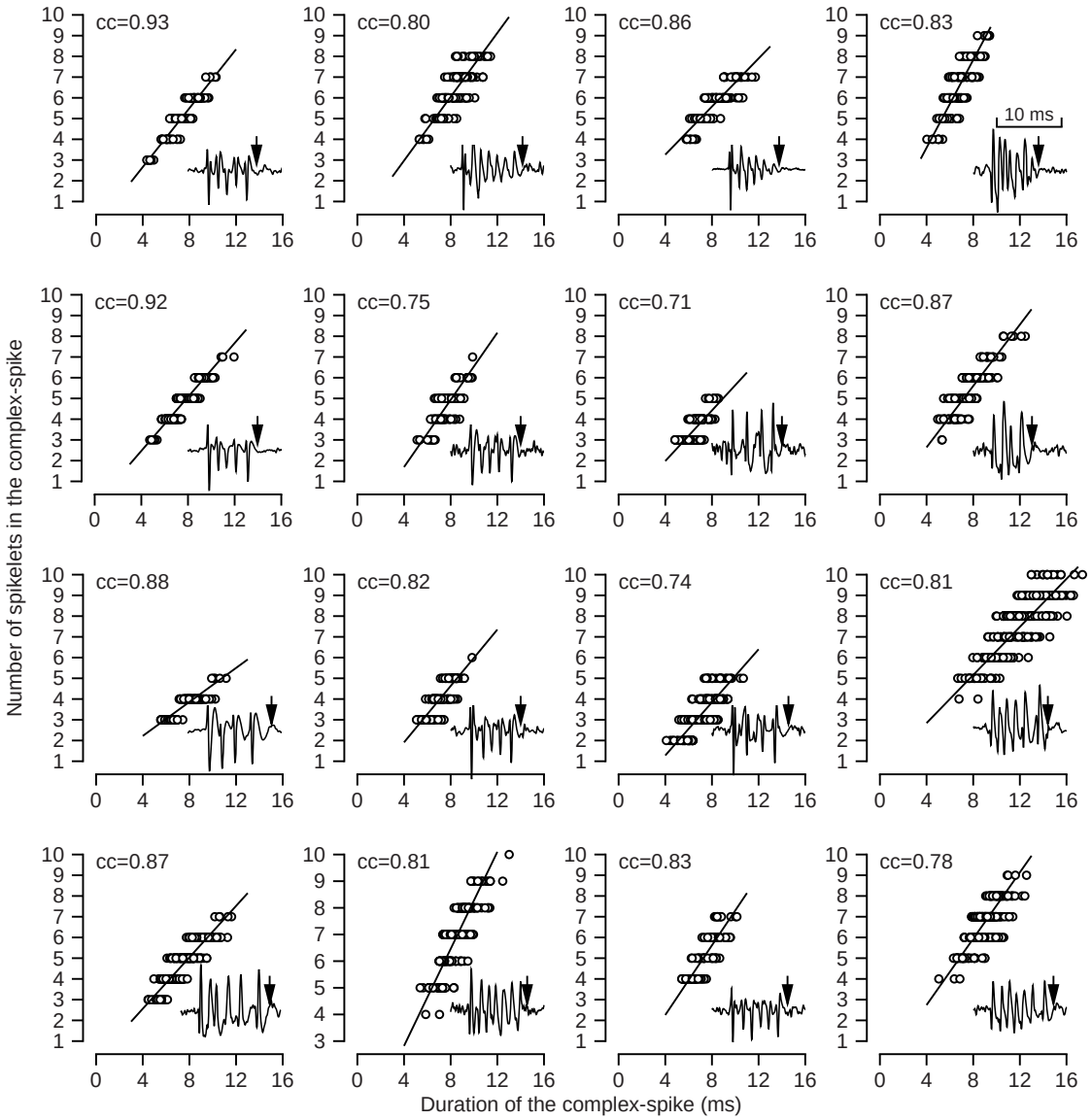
30. Maiz, J. *et al.* Prolonging the postcomplex spike pause speeds eyeblink conditioning. *Proc. Natl Acad. Sci. USA* **109,** 16726–16730 (2012).

©2014 **Macmillan Publishers Limited. All rights reserved**



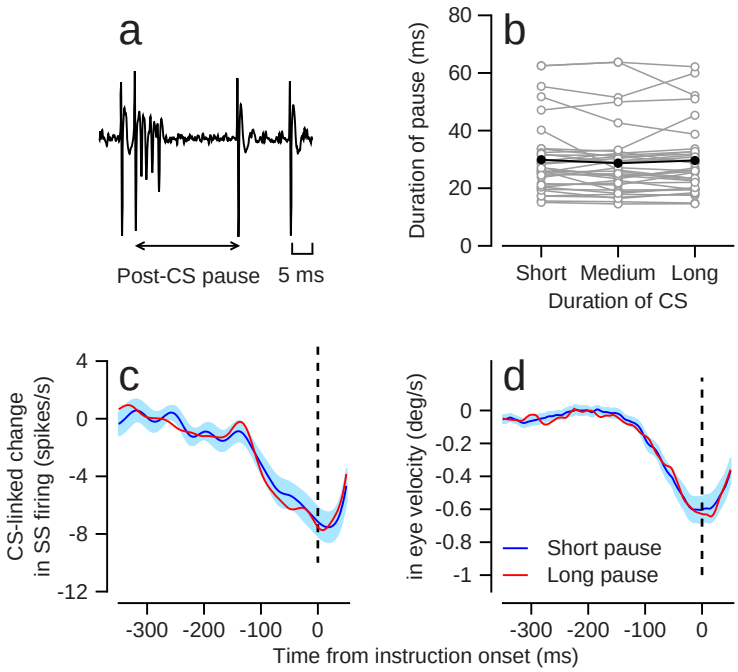
Extended Data Figure 1 | Direction tuning of floccular Purkinje cells during pursuit eye movements. **a**, Arrows show directions of ramp target motions used to assess direction tuning. Red and blue arrows show the on and off directions for simple-spike firing in the example Purkinje cell. **b**, **c**, Simple-spike firing rate and CS probability as a function of time during pursuit of vertical target motion in the on direction for simple-spike firing (**b**, red traces) and the off direction for simple-spike firing (**c**, blue traces). The target-position traces show the step-ramp target motion, which used a small step in one direction to obviate the need for saccades during the initiation of pursuit in the direction of the target ramp. **d**, **e**, Full direction tuning analysis for the simple-spike firing rate (**d**) and CS probability (**e**). Each graph comprises eight peri-stimulus time histograms that start 200 ms before the onset of target motion and are positioned appropriately for the direction of target motion. A polar plot appears at the centre of each panel showing the direction tuning as a black line, and the baseline response during fixation as a grey circle. In this example and in all Purkinje cells in our sample, simple-spike firing rate and CS probability had opposite preferred directions.

©2014 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 2 | Relationship between number of spikelets in a CS response and the duration of the extracellular potential. Each graph shows data for one of 16 Purkinje cells with CS responses that were isolated well enough to allow reliable counting of spikelets. Each symbol shows the measurements for a single CS response. The regression lines were obtained with a type-II regression analysis that gave equal weight to errors along the *x* and *y* axes. The duration of the CS response to a learning instruction did not vary as a function of tiny fluctuations in the magnitude of the instruction, caused by variation in the eye velocity at the time the instructive target motion occurred. The correlation coefficients for the individual graphs ranged from 0.71 to 0.93 and averaged 0.83. The arrows indicate the time chosen as the end of each CS waveform.

©2014 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 3 | Absence of relationship between duration of the post-CS pause in simple-spike firing rate and trial-over-trial depression or learning. **a**, An example CS waveform showing the measurement of the duration of the post-CS pause. **b**, Relationship between duration of pause and duration of CS. Each set of three grey symbols shows data from a single Purkinje cell. Black symbols show averages across the full sample. **c**, **d**, Trial-over-trial changes in simple-spike firing rate (**c**) and eye velocity (**d**). Red and blue traces show data for long versus short durations of post-CS pause. The light blue shaded area around the blue trace shows one standard error of the mean for the results with a short pause. Vertical dashed lines show the time of the instruction. We did not find any relationship between the magnitude of trial-over-trial depression of simple-spike firing or eye velocity and the duration of the post-CS pause in simple-spike responses. Thus, the post-CS pause might affect learning in the deep cerebellar nucleus[30], but does not seem to impact short-term learning in the cerebellar cortex. In terms of the methods, as before, we formed distributions of the duration of the post-CS pause for each Purkinje cell, trisected the distributions, and divided pairs of trials according to whether the post-CS pause in the instruction trial was long, medium or short (means: 42.6, 29.5 and 20.2 ms). The simple-spike activity was somewhat higher on trials with a short versus long post-CS pause (80.4 ± 7.0 versus 73.6 ± 6.9 s.e.m., two-tailed paired t-test, $P < 0.01$, $n = 34$).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13287

Human oocytes reprogram adult somatic nuclei of a type 1 diabetic to diploid pluripotent stem cells

Mitsutoshi Yamada[1]*, Bjarki Johannesson[1]*, Ido Sagi[2], Lisa Cole Burnett[3], Daniel H. Kort[4,5], Robert W. Prosser[4,5], Daniel Paull[1], Michael W. Nestor[1], Matthew Freeby[3], Ellen Greenberg[3], Robin S. Goland[3], Rudolph L. Leibel[3], Susan L. Solomon[1], Nissim Benvenisty[2], Mark V. Sauer[4,5] & Dieter Egli[1]

The transfer of somatic cell nuclei into oocytes can give rise to pluripotent stem cells that are consistently equivalent to embryonic stem cells[1–3], holding promise for autologous cell replacement therapy[4,5]. Although methods to induce pluripotent stem cells from somatic cells by transcription factors[6] are widely used in basic research, numerous differences between induced pluripotent stem cells and embryonic stem cells have been reported[7–11], potentially affecting their clinical use. Because of the therapeutic potential of diploid embryonic stem-cell lines derived from adult cells of diseased human subjects, we have systematically investigated the parameters affecting efficiency of blastocyst development and stem-cell derivation. Here we show that improvements to the oocyte activation protocol, including the use of both kinase and translation inhibitors, and cell culture in the presence of histone deacetylase inhibitors, promote development to the blastocyst stage. Developmental efficiency varied between oocyte donors, and was inversely related to the number of days of hormonal stimulation required for oocyte maturation, whereas the daily dose of gonadotropin or the total number of metaphase II oocytes retrieved did not affect developmental outcome. Because the use of concentrated Sendai virus for cell fusion induced an increase in intracellular calcium concentration, causing premature oocyte activation, we used diluted Sendai virus in calcium-free medium. Using this modified nuclear transfer protocol, we derived diploid pluripotent stem-cell lines from somatic cells of a newborn and, for the first time, an adult, a female with type 1 diabetes.

We have previously reported the derivation of pluripotent stem cells containing a reprogrammed genome derived from an adult somatic cell, and a haploid oocyte genome[12]. Development to the blastocyst stage only occurred in the presence of the oocyte genome; diploid nuclear transfer cells arrested development at the cleavage stages, failing to express embryonic genes. To improve developmental potential and transcriptional reprogramming in diploid nuclear transfer oocytes, we tested the effect of histone deacetylation (HDAC) inhibitors, as well as changes to the artificial activation protocol on developmental potential. These modifications were based on the report that HDAC inhibitors improved development after somatic cell nuclear transfer in mouse oocytes[13], and on our observation that parthenogenetic development was more efficient when oocytes were activated with the translation inhibitor puromycin[14] than when activated with the kinase inhibitor 6-dimethylaminopurine (6-DMAP)[12] (Extended Data Fig. 1). To minimize the effect of potential variation in oocyte quality, for each condition we used oocytes from at least four donors (Fig. 1a). We first tested the use of puromycin for oocyte activation in somatic cell nuclear transfer without removing the oocyte genome, resulting in efficient development to the blastocyst stage (Fig. 1a). However, development of diploid nuclear transfer oocytes still arrested at cleavage stages (Fig. 1a). Only when we applied the HDAC inhibitor scriptaid during the first embryonic interphase, did we observe development to the morula and blastocyst stages at a low frequency (Fig. 1a). Further improvement in developmental potential was observed when both puromycin and 6-DMAP were combined during oocyte activation, resulting in development to expanded blastocysts (Fig. 1a, b and Extended Data Fig. 2a, b). Puromycin promotes oocyte activation by inhibiting translation of cyclin B[15,16], whereas 6-DMAP inhibits the activity of meiotic kinases; their combined use may result in a more complete or more rapid inactivation of meiotic kinases. These results show that an improved activation protocol and the use of an HDAC inhibitor enabled development of nuclear transfer cells to the blastocyst stage in the absence of the oocyte genome.

Because development beyond the cleavage stage requires gene expression[17], development to the morula and blastocyst stages indicates transcriptional activity of the transferred somatic cell genome. Whereas previous nuclear transfer protocols did not result in expression of a green fluorescent protein (GFP) transgene contained in the somatic cell genome[12], 58% (14 of 24) of the nuclear transfer cells treated with HDAC inhibitor were GFP positive (Fig. 1b), and had a global gene expression profile similar to *in vitro* fertilized (IVF) embryos (Fig. 1c), demonstrating that transcriptional reprogramming was extensive. Of the 7 nuclear transfer blastocysts obtained using optimized protocols, 3 formed an outgrowth (Extended Data Fig. 2c), but none gave rise to an embryonic stem (ES) cell line.

More recently, it has become possible to derive diploid pluripotent stem-cell lines from fetal fibroblasts[18]. The derivation of cell lines from an 8-month-old subject with Leigh syndrome was also stated, although no karyotype or evidence of pluripotency was provided. While the use of the HDAC inhibitor to obtain blastocyst development is consistent with our data and a previous report[19], the authors also attributed successful derivation to the use of caffeine during oocyte enucleation to promote nuclear envelope breakdown and condensation of somatic chromatin, the use of a hormone stimulation protocol yielding a small number of high-quality oocytes, and to the use of an electrical pulse for oocyte activation. As it remained unclear whether diploid nuclear transfer ES cells could reliably be derived from postnatal somatic cells, we determined the requirements for blastocyst development and ES cell derivation using adult somatic cells of a type 1 diabetic female (age 32 years, age of onset 10 years) and human foreskin fibroblasts of a male newborn for nuclear transfer.

We first determined whether oocyte enucleation interfered with the condensation of transferred somatic chromatin, a process correlating with developmental potential[20]. When we transferred somatic cell genomes at G1 or G0 stages of the cell cycle into enucleated oocytes, 17 out of 23 (74%) assembled a spindle within 1–4 h after transfer as determined by microtubule birefringence[21] or immunostaining (Fig. 2a). Somatic chromosomes were condensed and phosphorylated on serine 28 of histone H3, but not aligned on a metaphase plate, because unreplicated

[1]The New York Stem Cell Foundation Research Institute, New York, New York 10032, USA. [2]Stem Cell Unit, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. [3]Naomi Berrie Diabetes Center, Department of Pediatrics, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA. [4]Center for Women's Reproductive Care, College of Physicians and Surgeons, Columbia University, New York 10019, USA. [5]Department of Obstetrics and Gynecology, College of Physicians and Surgeons, Columbia University, New York 10032, USA.

*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

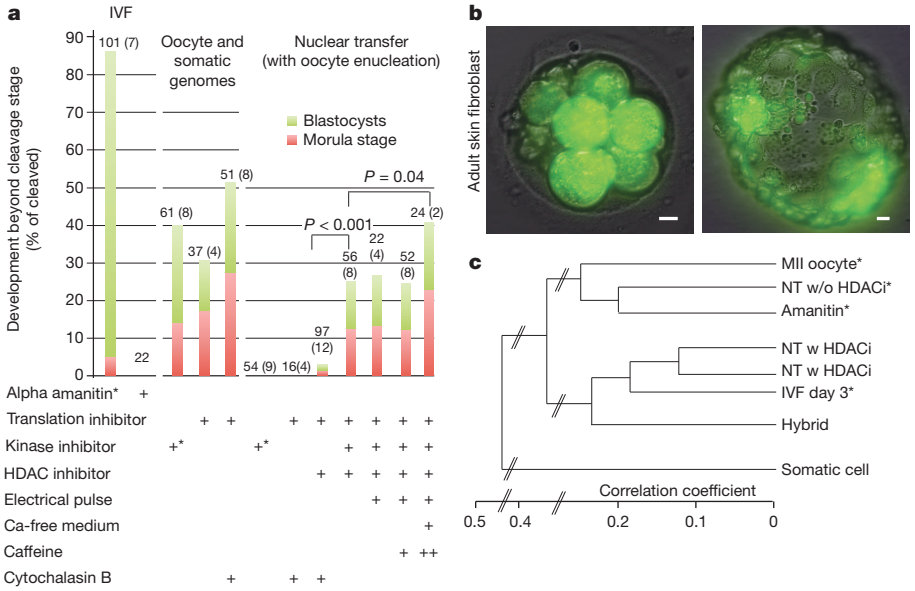


Figure 1 | Developmental potential of somatic cell nuclear transfer oocytes. **a**, Percentage of oocytes developing beyond the cleavage stage. The total number of oocytes and the number of oocyte donors (in parenthesis) contributing to a particular experiment is indicated above each column. ++, Caffeine was also added during oocyte transport. Statistical analysis was done using Chi-square test. **b**, Expression of a GFP transgene from the transferred somatic cell genome, at the cleavage stage and at the blastocyst stage. **c**, Cluster analysis of global gene expression profile after nuclear transfer of adult somatic cells, as well as oocytes and IVF embryos. *Data with an asterisk are from refs 12 and 27 for comparison. Scale bars, 10 μm. HDACi, HDAC inhibitor; NT, nuclear transfer.

chromosomes cannot form bipolar amphitelic attachments, as in a spindle of the MII oocyte (Fig. 2b). Although chromosome condensation did not occur in all oocytes receiving a somatic cell genome, this was not due to the enucleation procedure. Chromosome condensation and spindle assembly occurred with similar efficiency after transfer of a somatic cell genome into non-enucleated oocytes (10/13, 77%). Two non-enucleated oocytes that failed to assemble a spindle around somatic chromatin showed no phosphorylation of somatic histones (Fig. 2c), and segregation of the oocyte genome with the formation of a midbody positive for borealin (Fig. 2d), a component of the chromosome passenger complex localizing to the midbody at anaphase[22].

Because a rise in intracellular calcium concentration is a potent inducer of anaphase in oocytes, we investigated the effect of Sendai virus used for somatic cell fusion on intracellular calcium levels. Intact oocytes and karyoplasts—the by-product of oocyte enucleation—were equilibrated with the calcium indicator dye Fluo-4 and exposed to Sendai virus. Within minutes, an increase of Fluo-4 fluorescence was observed (Fig. 2e–h and Extended Data Fig. 3). Fluorescence was calcium-dependent, as oocytes incubated in calcium-free media showed decreased fluorescence compared to oocytes incubated in calcium-rich media (Extended Data Fig. 3). These results show that the fusion of somatic cells using Sendai virus can increase calcium influx and compromise the integrity of meiotic arrest. Therefore, nuclear transfer experiments from all but four oocyte donations were performed using Sendai virus diluted up to 20-fold. And in two donations, calcium was omitted from the medium used for nuclear transfer, as well as during incubation before oocyte activation. The absence of calcium in the medium was compatible with condensation of transferred somatic chromatin, resulting in efficient development to the blastocyst stage (Fig. 1a and Extended Data Fig. 4). Technical changes adapted from ref. 18, including an electrical pulse

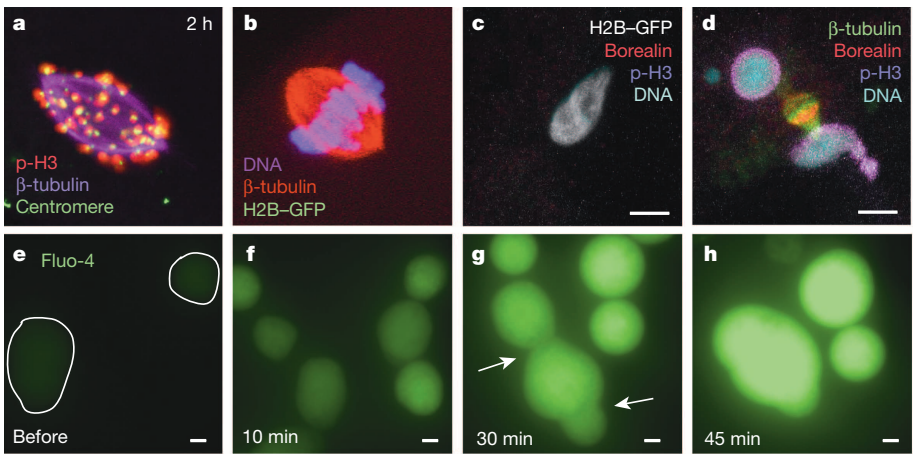


Figure 2 | Chromosome condensation and spindle assembly after somatic cell nuclear transfer. **a**, Spindle assembly on a somatic G1/G0 genome. Time indicates hours post transfer. p-H3, phosphorylated histone H3. **b**, Spindle of the human MII oocyte. **c**, Somatic nucleus transferred into the oocyte using undiluted Sendai virus. Note the lack of phosphorylated histone H3. **d**, Oocyte genome in the same egg. Note the segregation of oocyte chromosomes. **e–h**, Fluorescence time course of the calcium indicator dye Fluo-4 in oocyte karyoplasts before and after incubation with fusogenic Sendai virus (less than 20 s). Time after exposure is indicated. Arrows point to sites of fusion. Scale bar, 5 μm.

©2014 Macmillan Publishers Limited. All rights reserved

for oocyte activation and the use of caffeine during oocyte manipulation were most effective when combined with omitting calcium from the manipulation medium (Fig. 1a).

For the derivation of nuclear transfer ES (NT-ES) cell lines, we added fetal bovine serum (FBS) to the embryo culture and derivation media[18]. FBS promoted the formation of an inner cell mass at the expense of trophectoderm cells. Of eight nuclear transfer blastocysts generated without the addition of FBS, two contained exclusively (60 or more) trophectoderm cells (Extended Data Fig. 5a), and none gave rise to an ES cell line. In the presence of FBS, even blastocysts with a small number of cells formed a distinct inner cell mass (four out of four), with fewer than 20 trophectoderm cells (Extended Data Fig. 5a). Three out of four such blastocysts formed an outgrowth (Extended Data Fig. 5b) and developed into three cell lines containing a diploid male karyotype derived from foreskin fibroblasts (Extended Data Fig. 6). A NT-ES cell line with a diploid female karyotype was also derived from an adult somatic cell of a type 1 diabetic (Fig. 3a). All four cell lines expressed markers of pluripotency (Fig. 3b–d and Extended Data Fig. 6a–c), and lacked markers of the dermal fibroblasts (Fig. 3e). In an analysis of global gene expression patterns, NT-ES cells clustered closely with other human pluripotent stem cells, including ES cells from fertilized embryos and induced pluripotent stem (iPS) cells (Fig. 3f). In embryoid bodies and upon transplantation into immunocompromised mice, all four NT-ES cell lines gave rise to cell types of three germ layers (Fig. 3g and Extended Data Fig. 7). When exposing them to a combination of cellular patterning factors (see Methods), we found efficient differentiation into neurons (Fig. 3h), pancreatic and duodenal homeobox-1 (PDX1)-positive cells (Fig. 3i), and insulin-positive cells (Fig. 3j) that were able to secrete insulin into the medium upon potassium stimulation (Fig. 3k).

Although nuclear transfer blastocysts could be obtained with an efficiency of approximately 10%, developmental efficiency varied between different oocyte donors, even when other aspects of the nuclear transfer protocol were kept constant. To better understand the source of the variation in oocyte performance, we retrospectively assessed the effect of patient age and hormone stimulation protocol on developmental potential after nuclear transfer (Extended Data Fig. 8, 154 oocytes obtained from 18 donors contained in the last four columns of Fig. 1a). In contrast to a previous report[18], we found no association between the number of MII oocytes retrieved per donation and developmental competence (Extended Data Fig. 8b). In fact, the derivation of a NT-ES cell line from adult somatic cells was achieved using 9 oocytes selected randomly from a total of 31 MII oocytes obtained in a single donation. Oocytes from subjects age 21–26 years showed greater developmental potential compared to subjects age 27–32 years, consistent with data

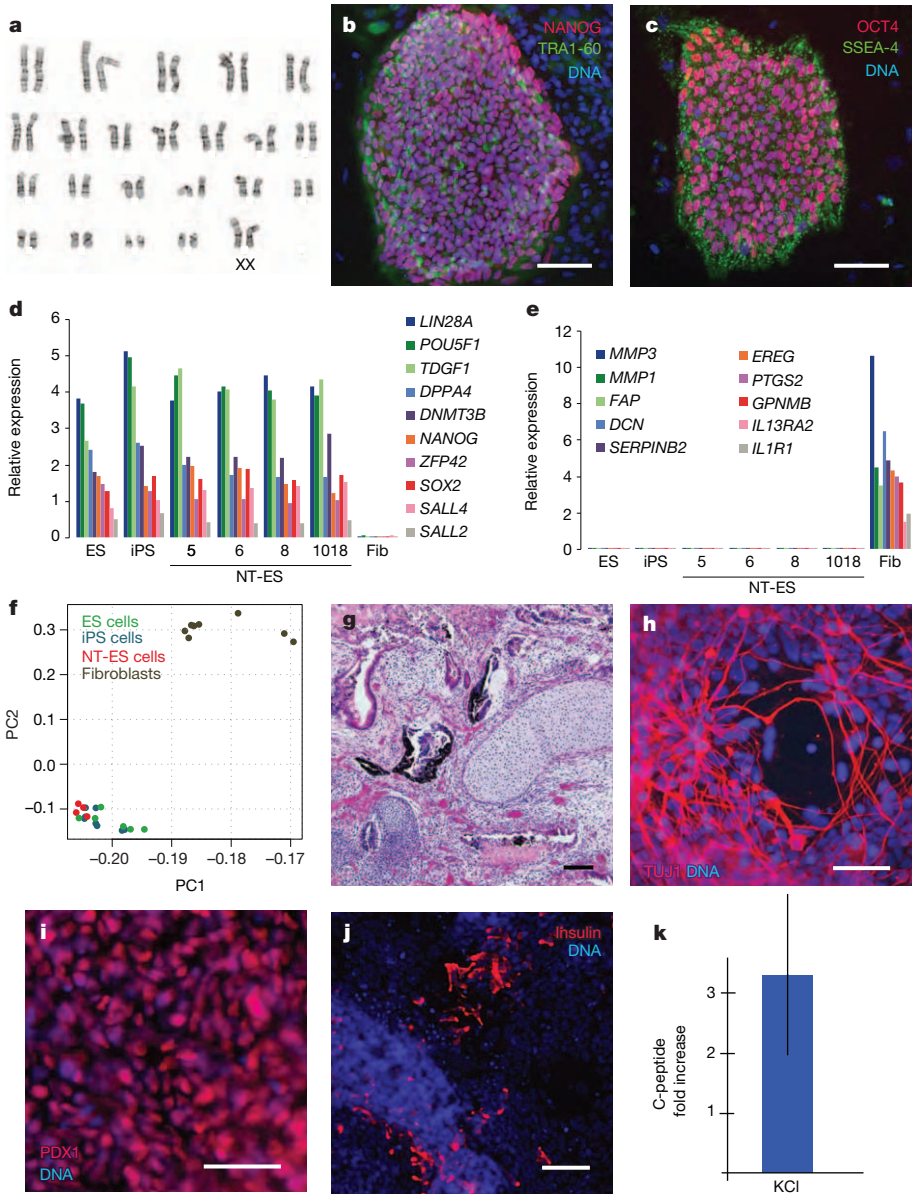


Figure 3 | Derivation of diploid NT-ES cells from adult somatic cells. **a–c**, **g–k**, Characterization of a NT-ES cell line derived from adult somatic cells of a female type 1 diabetic (ID 1018). **a**, Karyotype. **b**, **c**, Expression of pluripotency markers. **d**, **e**, Microarray analysis of NT-ES cell lines NT-ES5, NT-ES6, NT-ES8 (neonatal) and NT-ES 1018 (adult), and of fibroblasts (Fib), ES and iPS cells. **d**, Expression of pluripotency markers. **e**, Expression of fibroblast-specific genes. **f**, Principal component (PC) analysis of global gene expression patterns of NT-ES cells, ES cells from normally fertilized embryos, iPS cells and fibroblasts. Sample identity is provided in Extended Data Table 1. **g**, Teratoma analysis. **h**, Directed differentiation into neurons. **i**, **j**, Directed differentiation of NT-ES cells from a type 1 diabetic into pancreatic precursor cells (**i**) and insulin-producing cells (**j**). **k**, Insulin secretion upon stimulation with potassium. Bar indicates standard error of $n = 4$. Scale bar, 50 µm.

©2014 Macmillan Publishers Limited. All rights reserved

from clinical donor IVF cycles[23]. We observed no significant effect of the method of pituitary downregulation as gonadotropin releasing hormone (GnRH) antagonist and GnRH agonist protocols resulted in development to the blastocyst stage. We also observed no significant effect of the daily dose of gonadotropin used for stimulation. However, there was a trend towards a negative effect of the total duration (days) of gonadotropin stimulation required to reach a follicular size of 18 mm (Extended Data Fig. 8b). Although prolonged stimulation has been suspected to be a negative predictor of IVF outcome, clinical pregnancy rates have not been shown to be affected[24]. Thus, additional requirements for somatic cell reprogramming may reveal subtle biological differences that are not readily apparent in clinical IVF.

This retrospective analysis also revealed that one of the most relevant technical parameters for efficient blastocyst development was the use of diluted Sendai virus for fusion, whereas the use of FBS was most relevant for ES cell derivation (Extended Data Fig. 8c, d). Surprisingly, development to the blastocyst stage was equally efficient when either neonatal or adult fibroblasts were used for transfer.

Our results show that technical improvements enable development to the blastocyst stage after transfer of adult somatic cells. These changes obviated the requirement for the oocyte genome, although development with the oocyte genome remained more efficient (Fig. 1a). One interpretation of the beneficial effect of the oocyte genome is that it compensates for incomplete reprogramming of somatic chromatin. Future studies should lead to a better understanding of the molecular mechanisms of how these technical improvements affect reprogramming and developmental outcome.

In summary, we have demonstrated the derivation of human ES cells from neonatal and adult somatic cells by nuclear transfer. The stem cells are pluripotent and could be differentiated into insulin-producing beta cells, the cell type lost in patients with type 1 diabetes. These stem cells could therefore be used to generate cells for therapeutic cell replacement. Although it is now possible to induce stem-cell formation by overexpression of embryonic transcription factors[6], these cells are often differentiation-defective[7], contain aberrant patterns of cytosine methylation[9,10,25] and hydroxymethylation[8], acquire somatic coding mutations[26], and show biallelic expression of imprinted genes[11]. Studies comparing nuclear transfer ES cells to iPS cells of isogenic origin should enable evaluation of the quality of cells generated by different methods of reprogramming.

METHODS SUMMARY

Nuclear transfer was performed using somatic cells grown to confluence and cultured in 0.5% serum for 2–5 days before transfer to induce cell cycle arrest in G1/G0. Oocytes were enucleated either before or after transfer using microtubule birefringence[21] and somatic cells were fused to the oocyte using diluted Sendai virus, in calcium-free medium. Upon completion of the manipulation, oocytes were incubated for 45 min–1 h in calcium-free MCZB, then placed in calcium-containing medium for 10 min before activation. Oocyte activation was performed using electrical pulses, followed by 4 h incubation in medium containing 10 $\mu g\,ml^{-1}$ puromycin and 2 mM 6-DMAP as well as HDAC inhibitors scriptaid and nch51. HDAC inhibitor was retained for a total of 15 h during the first cell cycle. Culture to the blastocyst stage and ES cell derivation was performed containing 10%FBS tested for compatibility with growth of human ES cells. A detailed protocol for embryonic stem-cell derivation by somatic cell nuclear transfer into human oocytes is available on the *Nature* Protocol Exchange website (http://dx.doi.org/10.1038/protex.2014.013)[28].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 February; accepted 27 March 2014.
Published online 28 April 2014.

1. Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467,** 285–290 (2010).
2. Brambrink, T., Hochedlinger, K., Bell, G. & Jaenisch, R. ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc. Natl Acad. Sci. USA* **103,** 933–938 (2006).
3. Wakayama, S. *et al.* Equivalency of nuclear transfer-derived embryonic stem cells to those derived from fertilized mouse blastocysts. *Stem Cells* **24,** 2023–2033 (2006).
4. Tabar, V. *et al.* Therapeutic cloning in individual parkinsonian mice. *Nature Med.* **14,** 379–381 (2008).
5. Rideout, W. M. III, Hochedlinger, K., Kyba, M., Daley, G. Q. & Jaenisch, R. Correction of a genetic defect by nuclear transplantation and combined cell and gene therapy. *Cell* **109,** 17–27 (2002).
6. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
7. Koyanagi-Aoi, M. *et al.* Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc. Natl Acad. Sci. USA* **110,** 20569–20574 (2013).
8. Wang, T. *et al.* Subtelomeric hotspots of aberrant 5-hydroxymethylcytosine-mediated epigenetic modifications during reprogramming to pluripotency. *Nature Cell Biol.* **15,** 700–711 (2013).
9. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biol.* **13,** 541–549 (2011).
10. Ruiz, S. *et al.* Identification of a specific reprogramming-associated epigenetic signature in human induced pluripotent stem cells. *Proc. Natl Acad. Sci. USA* **109,** 16196–16201 (2012).
11. Pick, M. *et al.* Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells. *Stem Cells* **27,** 2686–2690 (2009).
12. Noggle, S. *et al.* Human oocytes reprogram somatic cells to a pluripotent state. *Nature* **478,** 70–75 (2011).
13. Van Thuan, N. *et al.* The histone deacetylase inhibitor scriptaid enhances nascent mRNA production and rescues full-term development in cloned inbred mice. *Reproduction* **138,** 309–317 (2009).
14. Paull, D. *et al.* Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* **493,** 632–637 (2013).
15. Winston, N. J. Stability of cyclin B protein during meiotic maturation and the first mitotic cell division in mouse oocytes. *Biol. Cell* **89,** 211–219 (1997).
16. Stern, S., Rayyis, A. & Kennedy, J. F. Incorporation of amino acids during maturation *in vitro* by the mouse oocyte: effect of puromycin on protein synthesis. *Biol. Reprod.* **7,** 341–346 (1972).
17. Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332,** 459–461 (1988).
18. Tachibana, M. *et al.* Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* **153,** 1228–1238 (2013).
19. Fan, Y. *et al.* Derivation of cloned human blastocysts by histone deacetylase inhibitor treatment after somatic cell nuclear transfer with beta-thalassemia fibroblasts. *Stem Cells Dev.* **20,** 1951–1959 (2011).
20. Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* **394,** 369–374 (1998).
21. Liu, L., Oldenbourg, R., Trimarchi, J. R. & Keefe, D. L. A reliable, noninvasive technique for spindle imaging and enucleation of mammalian oocytes. *Nature Biotechnol.* **18,** 223–225 (2000).
22. Gassmann, R. *et al.* Borealin: a novel chromosomal passenger required for stability of the bipolar mitotic spindle. *J. Cell Biol.* **166,** 179–191 (2004).
23. Cohen, M. A., Lindheim, S. R. & Sauer, M. V. Donor age is paramount to success in oocyte donation. *Hum. Reprod.* **14,** 2755–2758 (1999).
24. Bar-Hava, I. *et al.* Controlled ovarian hyperstimulation: does prolonged stimulation justify cancellation of *in vitro* fertilization cycles? *Gynecol. Endocrinol.* **21,** 232–234 (2005).
25. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).
26. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471,** 63–67 (2011).
27. Egli, D. *et al.* Reprogramming within hours following nuclear transfer into mouse but not human zygotes. *Nature Commun.* **2,** 488 (2011).
28. Egli, D. & Chia, G. A protocol for embryonic stem cell derivation by somatic cell nuclear transfer into human oocytes. *Protocol Exch.* http://dx.doi.org/10.1038/protex.2014.013 (2014).

Acknowledgements This research was supported the New York Stem Cell Foundation (NYSCF) and a New York State Stem Cell Science (NYSTEM) IIRP Award no. C026184, and the Russell Berrie Foundation Program in Cellular Therapies of Diabetes. We thank S. Mitalipov for helpful discussions and providing reagents, S. Micucci for counting cells in S-phase, and Z. Hall for critical reading of the manuscript. D.E. is a NYSCF-Robertson Investigator.

Author Contributions M.V.S. supervised the research oocyte donation program and retrieved oocytes. D.E. designed, performed and interpreted nuclear transfer experiments, derived ES cells with M.Y., and wrote the paper with input from all authors. M.Y. performed statistical analysis, M.Y. and B.J. performed stem cell characterization and differentiation, L.C.B. performed neuronal differentiation, I.S. and N.B. performed gene expression analysis, M.W.N. assisted with calcium experiments, D.H.K. performed data analysis, D.P. assisted in nuclear transfer experiments, R.W.P. collected developmental data of IVF embryos, M.F. made the skin biopsy, E.G. coordinated human subjects research, R.S.G. wrote the IRB protocol, R.L.L. contributed project planning, S.L.S. created the environment specifically for this work.

Author Information Microarray data are available at GEO under accession numbers GSE54849 and GSE54876. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E. (d.egli@nyscf.org).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13268

Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing

Volker Hovestadt[1]*, David T. W. Jones[2]*, Simone Picelli[1], Wei Wang[1], Marcel Kool[2], Paul A. Northcott[2], Marc Sultan[3], Katharina Stachurski[4], Marina Ryzhova[5], Hans-Jörg Warnatz[3], Meryem Ralser[3], Sonja Brun[6], Jens Bunt[7,8], Natalie Jäger[9], Kortine Kleinheinz[1,9], Serap Erkek[2,10], Ursula D. Weber[1], Cynthia C. Bartholomae[11,12], Christof von Kalle[11,12], Chris Lawerenz[13], Jürgen Eils[13], Jan Koster[8], Rogier Versteeg[8], Till Milde[14,15], Olaf Witt[14,15], Sabine Schmidt[16], Stephan Wolf[16], Torsten Pietsch[17], Stefan Rutkowski[18], Wolfram Scheurlen[19], Michael D. Taylor[20,21,22], Benedikt Brors[9], Jörg Felsberg[23,24], Guido Reifenberger[23,24], Arndt Borkhardt[4], Hans Lehrach[3], Robert J. Wechsler-Reya[6], Roland Eils[9,25,26,27], Marie-Laure Yaspo[3], Pablo Landgraf[4], Andrey Korshunov[28,29], Marc Zapatka[1], Bernhard Radlwimmer[1], Stefan M. Pfister[2,14] & Peter Lichter[1,27]

Epigenetic alterations, that is, disruption of DNA methylation and chromatin architecture, are now acknowledged as a universal feature of tumorigenesis[1]. Medulloblastoma, a clinically challenging, malignant childhood brain tumour, is no exception. Despite much progress from recent genomics studies, with recurrent changes identified in each of the four distinct tumour subgroups (WNT-pathway-activated, SHH-pathway-activated, and the less-well-characterized Group 3 and Group 4)[2–4], many cases still lack an obvious genetic driver. Here we present whole-genome bisulphite-sequencing data from thirty-four human and five murine tumours plus eight human and three murine normal controls, augmented with matched whole-genome, RNA and chromatin immunoprecipitation sequencing data. This comprehensive data set allowed us to decipher several features underlying the interplay between the genome, epigenome and transcriptome, and its effects on medulloblastoma pathophysiology. Most notable were highly prevalent regions of hypomethylation correlating with increased gene expression, extending tens of kilobases downstream of transcription start sites. Focal regions of low methylation linked to transcription-factor-binding sites shed light on differential transcriptional networks between subgroups, whereas increased methylation due to re-normalization of repressed chromatin in DNA methylation valleys was positively correlated with gene expression. Large, partially methylated domains affecting up to one-third of the genome showed increased mutation rates and gene silencing in a subgroup-specific fashion. Epigenetic alterations also affected novel medulloblastoma candidate genes (for example, *LIN28B*), resulting in alternative promoter usage and/or differential messenger RNA/microRNA expression. Analysis of mouse medulloblastoma and precursor-cell methylation demonstrated a somatic origin for many alterations. Our data provide insights into the epigenetic regulation of transcription and genome organization in medulloblastoma pathogenesis, which are probably also of importance in a wider developmental and disease context.

Medulloblastoma is an embryonal tumour with few differentiation markers and minimal normal-cell infiltration, making it well suited to epigenomic studies (where heterogeneity can be confounding). Thirty-four tumour and eight control samples were selected from a larger methylation array cohort for whole-genome bisulphite-sequencing analysis[5] (WGBS; average 28.5-fold average coverage and 91.4% of CpGs covered $\geq$10×; Supplementary Tables 1, 2 and Methods). Individual CpG sites showed a predominantly bimodal methylation pattern (either fully unmethylated or fully methylated), linked to CpG density (Extended Data Fig. 1a–c). Non-CpG methylation was low in tumours, but higher in adult cerebellum (Extended Data Fig. 1d, e), as recently reported for other brain regions[6].

Matching gene expression data enabled a comprehensive assessment of correlation with DNA methylation (see Methods), revealing negatively and positively correlating regions (CRs). Very few negative CRs were in promoter-associated CpG islands (CGIs). Of 13,380 genes with a promoter CGI, 333 (2.5%) were significantly differentially methylated, and only 79 showed negative correlation with expression (Extended Data Fig. 2a and Supplementary Table 3a). The *WNK2* tumour suppressor[7], for example, was silenced by CGI hypermethylation in most WNT medulloblastomas and some SHH and Group 3 samples (Extended Data Fig. 2b, c). Overall, medulloblastomas showed slightly higher CGI methylation than normal cerebellum, but without evidence for a CGI methylator phenotype (Extended Data Fig. 2d). Thus, the classical notion of gene silencing through promoter hypermethylation was not a prominent feature, and many differentially methylated promoters were at genes that were not expressed even in unmethylated samples.

[1]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [2]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [3]Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, Berlin 14195, Germany. [4]Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich Heine University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. [5]Department of Neuropathology, NN Burdenko Neurosurgical Institute, 4th Tverskaya-Yamskaya 16, Moscow 125047, Russia. [6]Tumor Initiation and Maintenance Program, National Cancer Institute (NCI)–Designated Cancer Center, Sanford-Burnham Medical Research Institute, 2880 Torrey Pines Scenic Drive, La Jolla, California 92037, USA. [7]Queensland Brain Institute, University of Queensland, QBI Building, St Lucia, Queensland 4072, Australia. [8]Department of Oncogenomics, AMC, University of Amsterdam, Meibergdreef 9, Amsterdam 1105 AZ, the Netherlands. [9]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [10]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg 69117, Germany. [11]Division of Translational Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [12]National Center for Tumor Diseases (NCT), Im Neuenheimer Feld 460, Heidelberg 69120, Germany. [13]Data Management Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [14]Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Im Neuenheimer Feld 430, Heidelberg 69120, Germany. [15]Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [16]Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [17]Department of Neuropathology, University of Bonn Medical Center, Sigmund-Freud-Strasse 25, Bonn 53105, Germany. [18]Department of Paediatric Haematology and Oncology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany. [19]Cnopf'sche Kinderklinik, Nürnberg Children's Hospital, St.-Johannis-Mühlgasse 19, Nürnberg 90419, Germany. [20]Program in Developmental and Stem Cell Biology, The Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [21]Division of Neurosurgery, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [22]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [23]Department of Neuropathology, Heinrich Heine University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. [24]German Consortium for Translational Cancer Research (DKTK), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [25]Institute of Pharmacy and Molecular Biotechnology (IPMB), University of Heidelberg, Heidelberg 69120, Germany. [26]Bioquant Center, University of Heidelberg, Im Neuenheimer Feld 267, Heidelberg 69120, Germany. [27]Heidelberg Center for Personalised Oncology (DKFZ-HIPO), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [28]Department of Neuropathology, University of Heidelberg, Im Neuenheimer Feld 220, Heidelberg 69120, Germany. [29]Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 220-221, Heidelberg, 69120 Germany.

*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

Much more abundant were negative CRs extending several kilobases (kb) downstream of the promoter (promoter downstream (pd)CRs), affecting 1,194 genes (Supplementary Table 3b). Correlation density peaked at ~2 kb downstream of the transcription start site (TSS), with most pdCRs extending much farther (often tens of kilobases, and sometimes greater than 50 kb; Fig. 1a–c). For example, *RUNX2*, encoding a WNT-activated transcription factor[8], showed a hypomethylated pdCR and high expression exclusively in WNT medulloblastomas (Fig. 1d, e). Overall, pdCRs were found in 8.4% of expressed genes, with many showing tumour-subgroup-specific differential methylation (Fig. 1f, Extended Data Fig. 3a, b and Supplementary Table 3b). Methylation and expression levels of genes harbouring a pdCR clearly distinguished between the tumour subgroups (Fig. 1g). Almost 20% of genes specifically expressed by a subgroup contain a pdCR (636 out of 3,304), suggesting that this methylation pattern plays an important part in determining the distinct transcriptomes of molecular variants of medulloblastoma (Supplementary Table 3b, c).

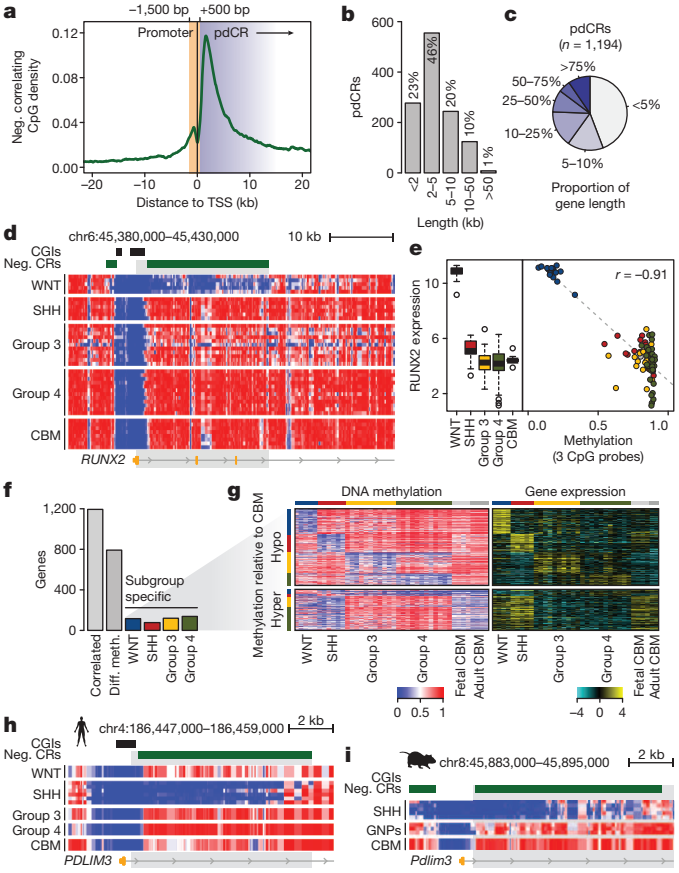


Figure 1 | Negative correlation between methylation and expression is enriched in extensive regions downstream of promoters. **a**, Density plot of CpGs negatively (neg.) correlating with gene expression with respect to distance from the TSS ($n = 36$ samples). **b**, **c**, pdCR length in kb (**b**) and as a proportion of gene length (**c**). **d**, Methylation plot for *RUNX2* ($n = 42$), showing a large WNT-medulloblastoma hypomethylated pdCR extending ~19 kb downstream of the promoter. Grey box, region of interest; red, methylated; blue, unmethylated. **e**, Correlation of expression and methylation for the *RUNX2* pdCR in the extended validation cohort ($n = 95$). CBM, cerebellum; *r*, Pearson's correlation coefficient. **f**, Overview of genes associated with a pdCR (correlated), those that are significantly differentially methylated (Diff. meth.) across medulloblastoma subgroups and controls, and those that are specifically differentially methylated in a given tumour subgroup. **g**, Heat-map representation of methylation ($n = 42$) and gene expression ($n = 36$) of subgroup-specific pdCRs. **h**, **i**, pdCR at *PDLIM3* in human ($n = 42$) (**h**) and murine ($n = 9$) (**i**) SHH medulloblastoma, but not in matched GNPs, indicating a somatic origin.

Analysis of publicly available WGBS data from various tissues (Supplementary Table 2d) revealed pdCRs as a universal feature, more prominent in homogeneous cell populations (for example, sorted neuronal or haematopoietic lineages versus infiltrative tumours or whole organs). A similar pattern in CpG 'shores' and 'shelves' (up to 2 kb and 2–4 kb from CGIs, respectively) has been described in some tissues using array-based methods or small WGBS cohorts, and has been linked to tissue-specific expression patterns[9–11] and to the presence of H3K4me3 (ref. 12), an active chromatin mark. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) in two medulloblastoma cell lines (D425 and MEB-MED-8A) confirmed an overlap of pdCRs and elevated H3K4me3 (Extended Data Fig. 3c, d). Our results highlight the extent of pdCRs as an abundant, previously underappreciated pattern of correlation between methylation and expression. Possible explanations for this hypomethylation range from passive methylation loss during cell division at highly RNA-polymerase-occupied regions (inaccessible to DNA methyltransferases), to a regulated process that actively marks genes for increased transcription.

We also produced WGBS profiles for murine SHH medulloblastomas arising owing to inactivation of patched homologue 1 (*Ptch1*; *Math1-creER^{T2}:Ptch1$^{fl/fl}$* (ref. 13)) together with matched purified cerebellar granule neuron precursors (GNPs)—the precursor for SHH medulloblastoma and therefore the optimal normal control. When comparing mouse tumours and precursor cells, 161 pdCRs larger than 2 kb were detected (Supplementary Table 3d). For example, *Pdlim3* shows a clear pdCR and high expression in human and mouse SHH medulloblastoma, but not in GNPs (Fig. 1h, i, and Extended Data Fig. 3e, f). Other known SHH medulloblastoma genes, such as *Ptch1*, *Cdk6* and *Boc*, showed a similar pattern. Thus, although some DNA methylation differences between the subgroups are probably linked to an epigenetic 'fingerprint' or 'memory' of developmental lineages, alterations also arise somatically during tumour development.

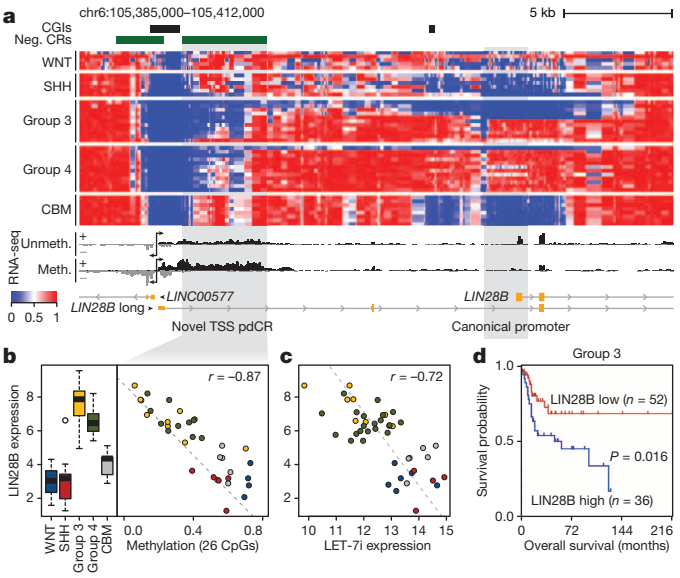


Figure 2 | Differential methylation around *LIN28B* reveals a novel promoter, tightly correlated with expression in Group 3 and Group 4 medulloblastomas. **a**, Methylation plot for the *LIN28B* promoter region ($n = 42$), with a novel TSS giving rise to an alternative transcript (*LIN28B* long). Arrows indicate bidirectional enhancer RNA transcription. Representative RNA-seq data are shown for samples methylated (Meth.) or unmethylated (Unmeth.) at the canonical promoter. Red, methylated; blue, unmethylated. CBM, cerebellum. **b**, Correlation of expression and methylation of an upstream CR at the novel TSS. *r*, Pearson's correlation coefficient. **c**, Correlation of *LIN28B* and *LET-7i* ($n = 59$). **d**, Association of *LIN28B* expression with overall survival in Group 3 medulloblastoma (Kaplan–Meier analysis, *P*, log-rank test). See also Extended Data Fig. 5f.

©2014 Macmillan Publishers Limited. All rights reserved

We next sought to identify novel promoters regulated by pdCRs. RNA sequencing (RNA-seq) data for 43 medulloblastomas and 7 control samples were assessed for splice junctions linking to unannotated 5′ exons (that is, novel first exons; see Methods). This revealed 2,479 exons not annotated in RefSeq in 1,937 genes (Supplementary Table 3e). Of these exons, 1,714 were listed in GENCODE, providing support for our method. Around 40% were upstream of the annotated 5′ end of the gene, with the remainder being internal. For example, *C9orf3*, hosting the miRNA cluster *MIR-23b–27b–24-1* cluster, was associated with a novel gene-body start site that harboured a pdCR specifically in the WNT subgroup (Extended Data Fig. 4a). Upregulation of this microRNA (miRNA) cluster has previously been noted in WNT medulloblastomas[14], and correlation with methylation was observed in our miRNA data set (Extended Data Fig. 4b–e).

To avoid confounding signals from canonical promoters, we focused on novel exons located >15 kb upstream of the previously annotated start site (the farthest being >500 kb upstream). Of 262 such exons, 49 showed patterns indicative of a pdCR (Supplementary Table 3d). One notable example is the miRNA-processing gene *LIN28B*, which showed differential regulation at a novel alternative first exon in Group 3 and Group 4 medulloblastomas. A pdCR at this start site (representing a bidirectional promoter with *LINC00577*) is hypomethylated in almost all Group 3 and Group 4 medulloblastomas, with increased LIN28B expression, whereas the canonical promoter was mostly hypermethylated (Fig. 2a, b and Extended Data Fig. 5a–c). *LIN28B* regulates multiple oncogenic processes, in part by downregulating the tumour-suppressive *LET-7* miRNA family[15]. Group 3 and 4 medulloblastomas showed significantly lower expression of most *LET-7* miRNAs compared with WNT and SHH medulloblastomas (Fig. 2c and Extended Data Fig. 5d, e). High *LIN28B* expression has been linked with poor prognosis in neuroblastoma[16], and this also held true for Group 3 and 4 medulloblastomas (Fig. 2d and Extended Data Fig. 5f).

Broader inspection of genome-wide methylation revealed megabase-scale partially methylated domains (PMDs)[10,17–19], affecting up to one-third of the tumour genome (Fig. 3a, b and Extended Data Fig. 6a, b). These regions account for the decreased global methylation observed in WNT and Group 3 tumours (Extended Data Fig. 1c). Although PMDs have been noted as a prominent feature of tumour methylomes[1], with global hypomethylation common in cancer, the affected regions showed clear subgroup specificity in WNT and Group 3 medulloblastomas, and they were virtually absent in SHH and Group 4 tumours (Fig. 3b and Extended Data Fig. 6c–e).

PMDs have been linked with inactive chromatin and large-scale gene silencing[17,18,20]. Genes inside subgroup-specific PMDs were significantly less expressed than non-PMD genes in that subgroup (positive correlation with methylation; Supplementary Table 4a), even (to a lesser extent) in tumours without clear hypomethylation (Extended Data Fig. 7a, b). Intra-subgroup heterogeneity, however, suggested some divergence in PMD formation (for example, for calmin (*CLMN*) in Group 3; Extended Data Fig. 7c, d). ChIP-seq in D425 cells confirmed an association of PMDs with long, organized domains of repressive marks (H3K9me3, H3K27me3)[17–19], and an absence of the active H3K36me3 mark (Fig. 3c, d and Extended Data Fig. 7e). Longer PMDs showed widespread H3K9me3 bordered by H3K27me3, whereas shorter PMDs showed H3K27me3 only (Fig. 3d). A significantly increased somatic mutation rate was observed

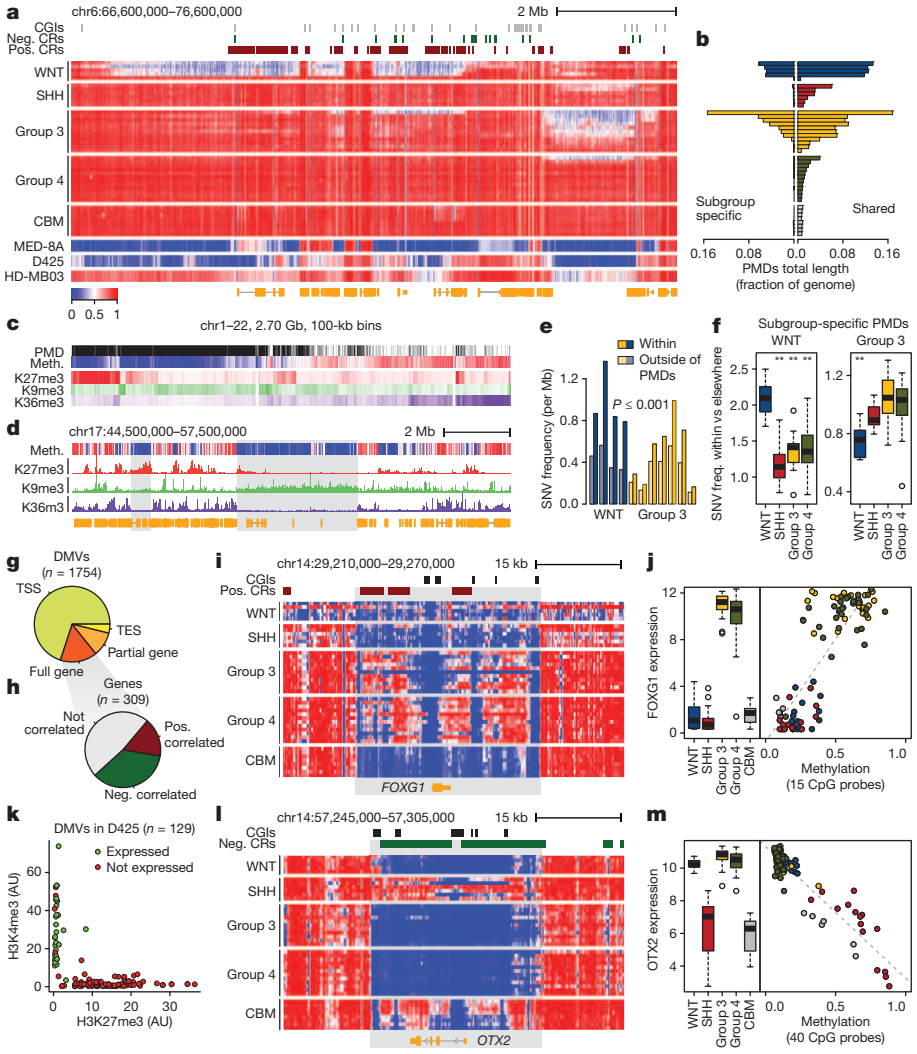


Figure 3 | Megabase-scale silenced domains and smaller DMVs are associated with distinct histone marks and positive correlation with expression. **a**, Example of a region on chromosome 6 showing large, subgroup-specific PMDs ($n = 42$). Negative (Neg.) and positive (Pos.) CRs are indicated. Partial methylation on chromosome 6 in WNT medulloblastomas (4/5 show monosomy 6) supports the notion that PMDs are stochastically acquired and not allele specific. Red, methylated; blue, unmethylated. CBM, cerebellum. **b**, Genome-wide quantification of PMDs. **c**, **d**, D425 cell line methylation (Meth.) and ChIP-seq data. PMDs are associated with increased H3K9me3 (for larger stretches) or H3K27me3 (typically smaller regions), and H3K36me3 depletion. **e**, Regions of the genome covered by a PMD show a higher somatic mutation frequency than non-PMD regions in a given sample (paired *t*-test). **f**, For WNT- or Group-3-specific PMDs, the ratio of single nucleotide variation (SNV) frequencies (freq.) within PMDs compared with elsewhere is higher in the PMD-containing subgroup. $^{**}P < 0.01$ (*t*-test). **g**, **h**, Most DMVs overlap gene regions, including many that encompass entire genes. Both negative and positively correlating DMVs are observed. **i**, **j**, Methylation plot around *FOXG1* ($n = 42$), which is encompassed by a positively correlating DMV. Re-establishment of methylation in Group 3 and Group 4 tumours correlates with increased expression ($n = 95$). **k**, DMVs in the D425 cell line show either the active H3K4m3 mark (when expressed) or the inactive H3K27me3 mark (when silenced), but rarely both. AU, arbitrary units. **l**, **m**, Methylation plot around *OTX2*, with a negatively correlating DMV, showing extensive hypomethylation in WNT, Group 3 and Group 4 tumours ($n = 42$) linked with higher expression in these groups ($n = 95$).

©2014 Macmillan Publishers Limited. All rights reserved

in PMDs (Fig. 3e, f), consistent with increased mutagenesis in late-replicating heterochromatin[21]. Mutant allele fractions indicated that most of this additional mutational burden is probably occurring early in tumorigenesis (data not shown).

The subgroup specificity of PMDs implies a non-random heterochromatization. Expression of tumour PMD-related genes was lower in normal cerebellum than non-PMD genes (Extended Data Fig. 7a, b), indicating pre-existing silencing in certain normal precursors. Variable hypomethylation within tumours, however, suggests a stochastic loss among different cell populations (Extended Data Fig. 6c, d). WGBS data of medulloblastoma cell lines support a gradual acquisition model, as hypomethylation was much more pronounced in long-established lines (MEB-MED-8A, D425) compared with a recent model (HD-MB03; Extended Data Fig. 6a). Thus, PMDs probably arise progressively due to failed transmission of methylation in late replicating, repressive domains[22], and are probably secondary to, rather than causative of, heterochromatin formation.

More localized hypomethylated regions surrounding smaller genes, termed DNA methylation valleys (DMVs) or lowly methylated 'canyons', have recently been reported as regulating tissue-specific gene expression[23,24]. This pattern, associated with H3K27me3 and/or H3K4me3 histone marks, is reminiscent of previously described, evolutionarily conserved polycomb-binding sites[25]. Our WGBS data revealed several hundred valleys per sample, with 1,754 unique DMVs overlapping a gene (Fig. 3g). Almost 20% ($n = 309$) encompassed an entire gene, and many were correlated with expression (Fig. 3h, Extended Data Fig. 8a–c and Supplementary Table 4b). As previously noted[24], a large number of DMVs affect developmental transcription factors and neuronal lineage genes.

Interestingly, 16.8% of DMVs showed a positive correlation between methylation and expression. These regions are reminiscent of PMDs, but are present in all tumours and controls, are much shorter, and have a higher CpG density (Supplementary Table 4c). One example is the region around *FOXG1*, a transcription factor linked with stem cell self-renewal in Group 3 and Group 4 medulloblastomas[26]. This gene is hypomethylated and not expressed in normal cerebellum, whereas hypermethylation of the DMV in Group 3 and Group 4 tumours is coupled with markedly increased expression (Fig. 3i, j).

ChIP-seq data for DMVs in D425 cells that are also present in primary tumours and encompass an entire gene ($n = 129$) revealed a notable pattern of either H3K4me3 (when the gene was expressed) or H3K27me3 (no expression) marks, with few bivalent sites (Fig. 3k). The absence of DNA methylation in DMVs showing positive correlation between methylation and expression may reflect a particularly dense heterochromatin that is inaccessible to the methylation machinery. Increased methylation is associated with gain of the active mark H3K4me3, loss of H3K27me3, and increased expression (for example, at the *PITX2* locus; Extended Data Fig. 8d, e). Thus, reactivation may involve a 'normalization' of the chromatin state from densely packed heterochromatin to a more open conformation, with re-establishment of the DNA methylation pattern typical of other genomic regions. DMVs are generally thought to be hypomethylated in normal tissues, and have been reported to contain CGIs that are particularly prone to silencing by hypermethylation in cancer[24,27,28]. We demonstrate, however, that cancer-specific gain of DNA methylation in DMVs is also linked to reactivation of repressed genes. This relationship is similar to that observed previously for a subset of PMDs in embryonic stem cells, which showed high levels of H3K27me3 and low methylation presumably linked with polycomb-mediated repression. Decreased H3K27me3 and increased DNA methylation in several of these PMDs, indicating relieved repression, was noted in differentiated IMR90 fibroblasts[19]. Both shorter PMDs and positively correlating DMVs may therefore be linked to similar regulatory processes.

Negatively correlating DMVs ($n = 111$) were also identified, with hypomethylation and concomitant upregulation of expression (Supplementary Table 4d). One notable example was *OTX2*, a known medulloblastoma oncogene with roles in cell cycle regulation and differentiation[29]. It is strongly hypomethylated and markedly overexpressed in WNT, Group 3 and Group 4 tumours (Fig. 3l, m).

OTX2 was also highlighted by an analysis of additional focal regions that did not fit into the categories described earlier, but clearly showed differential methylation. Localized lowly methylated regions (LMRs) have previously been linked with transcription-factor-binding sites and gene regulation[19,30]. To prove a functional link between LMRs and transcription-factor binding in medulloblastoma, we performed ChIP-seq for OTX2. As expected, ChIP-seq peaks were associated with hypomethylation, which was strongest at the centre of the peak. The periodicity in flanking methylation levels matched the size of a nucleosome, suggesting that hypomethylation of transcription-factor-binding sites may spread to adjacent linker regions (Fig. 4a). Conversely, 75.9% of LMRs that contained an OTX2-binding motif, identified in a systematic analysis, overlapped with an OTX2 ChIP-seq peak (Extended Data Fig. 9a).

Approximately 70,000 LMRs were identified per sample, with >45,000 unique regions showing subgroup-specific patterns (Supplementary Table 5a). Clustering revealed six groups: four with specificity for each medulloblastoma subgroup, plus one non-SHH/non-cerebellum and one non-Group 3/non-Group 4 cluster (Fig. 4b and Extended Data Fig. 9b).

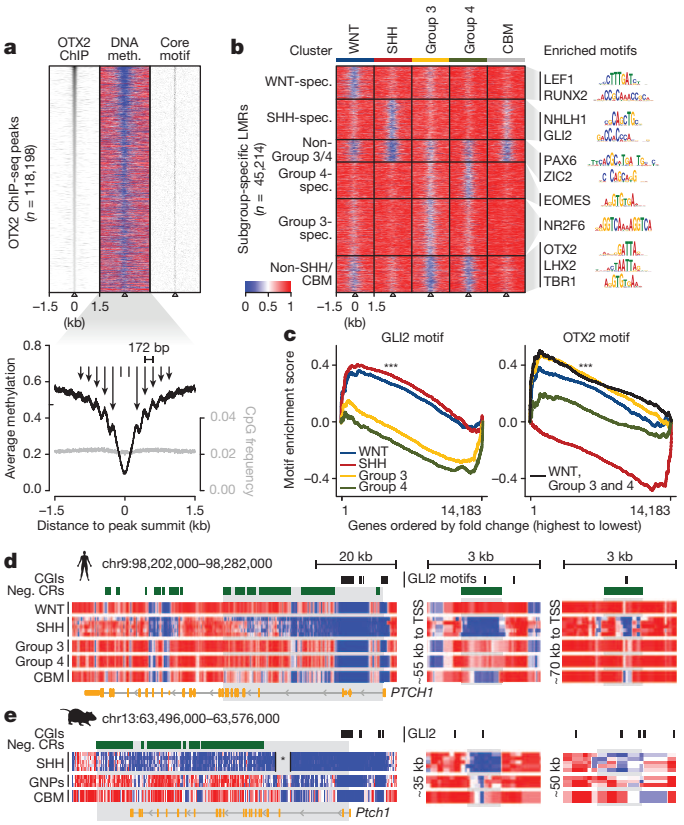


Figure 4 | Focal LMRs mark binding sites for key transcriptional regulators. **a**, Sites of OTX2 binding as defined by ChIP-seq in the D425 cell line are marked by hypomethylation and possess an OTX2-binding motif. Methylation (meth.) around the peak summit reveals a nucleosome periodicity. Red, methylated; blue, unmethylated. **b**, Heat-map representation of subgroup-specific (spec.) LMRs (±1.5 kb around centre) grouped according to *k*-means clustering. Motif analysis on these LMR clusters revealed enrichment for several key transcription factors, with representative examples shown (see Supplementary Table 5b). CBM, cerebellum. **c**, Examples of downstream regulation of gene expression. Binding motifs for the identified transcription factors are strongly enriched in the LMRs of the most highly overexpressed genes in the respective subgroup. ***$P < 0.001$ (motif enrichment score, see Methods). **d**, **e**, Example of upstream LMRs at the *PTCH1/Ptch1* locus in human/murine medulloblastomas with GLI2-binding sites. The same sites are not hypomethylated in mouse GNPs, indicating a somatic origin. There is also a pdCR at this site in both mouse ($n = 9$) and human ($n = 41$; sample ICGC_MB1, with homozygous *PTCH1* deletion, is not shown). Asterisk indicates deletion of *Ptch1* exon 3 (flanked by LoxP sites in the *Ptch1*$^{fl/fl}$ model).

©2014 Macmillan Publishers Limited. All rights reserved

In a systematic analysis, each cluster showed enrichment for certain transcription-factor-binding motifs, implicating known medulloblastoma-related genes and new candidates (for example, *NR2F6* in Group 3 medulloblastoma; Supplementary Table 5b). To test whether the transcription factors highlighted by LMR clustering might regulate downstream networks, we also looked at LMRs associated with the most highly expressed genes within the subgroups. Many transcription-factor-binding motifs identified in a particular LMR cluster were also highly enriched in these genes, demonstrating a key role for these transcription factors in determining subgroup identity (Fig. 4c and Supplementary Table 5b).

Analysis of differentially methylated LMRs therefore represents a powerful tool to provide a global, unbiased assessment of the key transcription factor regulatory networks acting in different tissues.

Analysis of our own and public mouse WGBS data again supported a somatic origin for many LMRs. *PTCH1/Ptch1*, for example, harboured several upstream LMRs with GLI2-binding sites in human and mouse SHH medulloblastoma that were not seen in normal cerebellum or GNPs (Fig. 4d, e and Extended Data Fig. 10). The GLI2-binding motif was enriched in mouse-tumour-specific LMRs, in keeping with the human data (Extended Data Fig. 9c and Supplementary Table 5c).

Our integrative analysis has uncovered a wealth of alterations in the methylome landscape of medulloblastoma. The regulatory effect of such epigenetic alterations in modulating the cancer phenotype allows for gradual global changes during tumour evolution, and their interplay with single time-point events such as genetic mutations will be an important future avenue of investigation. New targets that may be important for the pathophysiology of these tumours and for therapeutic intervention were also identified. These data, and the analysis methods applied, represent a valuable resource for further investigation into epigenetic regulatory mechanisms in cancer biology and in a wider context of other disease processes and organismal development.

METHODS SUMMARY

All patients contributing material to this study provided informed consent. Tumour samples were collected before adjuvant chemo- or radiotherapy. Control cerebellum samples were from a commercial supplier (BioChain). Cohort details are given in Supplementary Table 1. Tumour subgrouping was based on DNA methylation profiling, as previously described[5]. Genome-wide methylome profiles were derived by whole-genome bisulphite sequencing, as described in Methods. Sequencing metrics are given in Supplementary Table 2. CRs were derived by merging overlapping 5-CpG windows that showed correlation between DNA methylation and gene expression. Differential methylation and expression were assessed using analysis of variance and post-hoc testing. Further details are provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 July 2013; accepted 20 March 2014.
Published online 18 May 2014.

1. Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153,** 38–55 (2013).
2. Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* **123,** 465–472 (2012).
3. Jones, D. T. W. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488,** 100–105 (2012).
4. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nature Rev. Cancer* **12,** 818–834 (2012).
5. Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125,** 913–916 (2013).
6. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341,** 1237905 (2013).
7. Hong, C. *et al.* Epigenome scans and cancer genome sequencing converge on WNK2, a kinase-independent suppressor of cell growth. *Proc. Natl Acad. Sci. USA* **104,** 10974–10979 (2007).
8. Gaur, T. *et al.* Canonical WNT signaling promotes osteogenesis by directly stimulating *Runx2* gene expression. *J. Biol. Chem.* **280,** 33132–33140 (2005).
9. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.* **41,** 1350–1353 (2009).
10. Hansen, K. D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nature Genet.* **43,** 768–775 (2011).
11. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41,** 178–186 (2009).
12. Hodges, E. *et al.* Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* **44,** 17–28 (2011).
13. Yang, Z. J. *et al.* Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer Cell* **14,** 135–145 (2008).
14. Cho, Y. J. *et al.* Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J. Clin. Oncol.* **29,** 1424–1430 (2011).
15. Shyh-Chang, N. & Daley, G. Q. Lin28: primal regulator of growth and metabolism in stem cells. *Cell Stem Cell* **12,** 395–406 (2013).
16. Molenaar, J. J. *et al.* LIN28B induces neuroblastoma and enhances MYCN levels via *let-7* suppression. *Nature Genet.* **44,** 1199–1206 (2012).
17. Berman, B. P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature Genet.* **44,** 40–46 (2012).
18. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22,** 246–258 (2012).
19. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
20. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genet.* **41,** 246–250 (2009).
21. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488,** 504–507 (2012).
22. Aran, D., Toperoff, G., Rosenberg, M. & Hellman, A. Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* **20,** 670–680 (2011).
23. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nature Genet.* **46,** 17–23 (2014).
24. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153,** 1134–1148 (2013).
25. Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl Acad. Sci. USA* **104,** 5521–5526 (2007).
26. Manoranjan, B. *et al.* FoxG1 interacts with Bmi1 to regulate self-renewal and tumorigenicity of medulloblastoma stem cells. *Stem Cells* **31,** 1266–1277 (2013).
27. Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genet.* **39,** 237–242 (2007).
28. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nature Genet.* **39,** 232–236 (2007).
29. Bunt, J. *et al.* OTX2 directly activates cell cycle genes and inhibits differentiation in medulloblastoma cells. *Int. J. Cancer* **131,** E21–E32 (2012).
30. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the ICGC PedBrain Tumor Project, the German Cancer Research Center (DKFZ) Genomics and Proteomics Core Facility, the European Molecular Biology Laboratory (EMBL) Genomics Core Facility, M. Schick, R. Fischer, M. Bewerunge-Hudler, M. Knopf, R. Kabbe, A. Benner, R. Volckman and P. van Sluis for technical support and helpful discussion. Active Motif, Inc. is acknowledged for ChIP and library preparation. We also thank C. Plass for critical reading of the manuscript. This work was principally supported by the PedBrain Tumor Project contributing to the International Cancer Genome Consortium, funded by German Cancer Aid (109252) and the German Federal Ministry of Education and Research (BMBF, grants #01KU1201A, MedSys #0315416C and NGFN^{plus} #01GS0883). Additional support came from the DKFZ-Heidelberg Center for Personalized Oncology (DKFZ-HIPO), the Dutch Cancer Foundations KWF (2010-4713) and KIKA (M.Ko.), and the German Research Foundation (DFG; grant LA2983/2-1 to P.La.).

Author Contributions D.T.W.J., S.Pi., W.W., M.S., S.B., J.B., C.C.B., C.v.K., R.V., S.S., S.W., J.F. and P.La performed and/or coordinated experimental work. V.H., M.K., P.A.N., K.S., N.J., H.-J.W., M.Ra., K.K., S.E., C.L., J.E., J.K. and P.La. performed data analysis. M.Ry., T.M., O.W., T.P., S.R., W.S., M.D.T. and A.K. collected data and provided patient materials. V.H., D.T.W.J., M.K., P.A.N., M.Z., B.R., S.M.P. and P.Li. prepared the initial manuscript and figures. D.T.W.J., U.D.W., B.B., G.R., A.B., H.L., R.J.W.-R., R.E., M.-L.Y., A.K., P.La., M.Z., B.R., S.M.P. and P.Li. provided project leadership.

Author Information Short-read sequencing data have been deposited in the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession number EGAS00001000561. Methylation array data have been deposited in the Gene Expression Omnibus under accession number GSE54880. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.R. (b.radlwimmer@dkfz-heidelberg.de), S.M.P. (s.pfister@dkfz-heidelberg.de), or P.Li. (peter.lichter@dkfz-heidelberg.de).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13270

Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase

Anila K. Madiraju[1,2,3], Derek M. Erion[1,2,3], Yasmeen Rahimi[1], Xian-Man Zhang[1], Demetrios T. Braddock[4], Ronald A. Albright[4], Brett J. Prigaro[5], John L. Wood[6], Sanjay Bhanot[7], Michael J. MacDonald[8], Michael J. Jurczak[1], Joao-Paulo Camporez[1], Hui-Young Lee[1], Gary W. Cline[1], Varman T. Samuel[1], Richard G. Kibbey[1,2] & Gerald I. Shulman[1,2,3,9]

Metformin is considered to be one of the most effective therapeutics for treating type 2 diabetes because it specifically reduces hepatic gluconeogenesis without increasing insulin secretion, inducing weight gain or posing a risk of hypoglycaemia[1,2]. For over half a century, this agent has been prescribed to patients with type 2 diabetes worldwide, yet the underlying mechanism by which metformin inhibits hepatic gluconeogenesis remains unknown. Here we show that metformin non-competitively inhibits the redox shuttle enzyme mitochondrial glycerophosphate dehydrogenase, resulting in an altered hepatocellular redox state, reduced conversion of lactate and glycerol to glucose, and decreased hepatic gluconeogenesis. Acute and chronic low-dose metformin treatment effectively reduced endogenous glucose production, while increasing cytosolic redox and decreasing mitochondrial redox states. Antisense oligonucleotide knockdown of hepatic mitochondrial glycerophosphate dehydrogenase in rats resulted in a phenotype akin to chronic metformin treatment, and abrogated metformin-mediated increases in cytosolic redox state, decreases in plasma glucose concentrations, and inhibition of endogenous glucose production. These findings were replicated in whole-body mitochondrial glycerophosphate dehydrogenase knockout mice. These results have significant implications for understanding the mechanism of metformin's blood glucose lowering effects and provide a new therapeutic target for type 2 diabetes.

Initial investigations into metformin action found that this compound is a complex I inhibitor at millimolar concentrations[3,4]. More recent studies suggested that metformin activates AMP-activated protein kinase (AMPK) through decreases in hepatic energy charge (increasing [AMP]:[ADP] and/or [ADP]:[ATP] concentration ratios)[5,6] or through the upstream AMPK kinase, LKB1, leading to reduction of gluconeogenic gene transcription[6–8]. This effect may, however, be due to sensitization of gluconeogenic transcription to insulin through AMPK-mediated decreases in hepatic lipid content[9,10].

In contrast to these findings, it has been suggested that metformin suppresses gluconeogenesis independently of AMPK, instead altering hepatic energy charge[11] and inducing allosteric inhibition of glycolytic enzymes or adenylate cyclase and glucagon-activated gluconeogenic transcription[12]. However, these hypotheses are inconsistent with findings that metformin does not alter hepatic adenine-nucleotide levels[13].

Given these conflicting results, it is evident that the mechanisms by which guanide/biguanides exert their therapeutic effects remain to be explained. Variability in data on metformin action may arise because of differences in the dose of metformin administered *in vivo* ranging from 50 to 500 mg kg^{-1}, as well as observation of chronic versus acute responses to the drug.

In tracing the development of this drug class to the rapid-acting parent compound galegine (a mono-guanide), we considered that the acute glucose lowering effects reported might provide a model in which to investigate potential targets of action of the entire class. Within 20 min of intravenous infusion of galegine, plasma glucose and insulin concentrations decreased (Extended Data Fig. 1a, b), and plasma lactate concentrations increased eightfold (Extended Data Fig. 1c), independently of any changes in hepatic gluconeogenic gene expression (Extended Data Fig. 1d). Galegine also increased total AMPKα2 activity (Extended Data Fig. 1e). However, intravenous administration of the AMPK activator A-769662 failed to decrease plasma glucose concentrations or endogenous glucose production (EGP) (Extended Data Fig. 1f, g) despite comparable increases in AMPK activity 20 min after infusion (Extended Data Fig. 1h). Thus although guanide/biguanide treatment may activate AMPK, acute AMPK activation is not sufficient to reduce EGP.

In considering other mechanisms for metformin action, we were struck by the consistent observation of a marked increase in plasma lactate concentrations in guanide/biguanide-treated animals. This increase could not be attributed to reduced activity of enzymes regulating pyruvate metabolism (Extended Data Fig. 2a), as metformin, galegine and phenformin had no effect on pyruvate carboxylase, citrate synthase and alanine aminotransferase activity (Extended Data Fig. 2b–d). We developed the alternative hypothesis that the striking increase in plasma lactate concentrations may be due to metformin-induced increases in the cytosolic redox state, known to manifest as an increased [lactate]:[pyruvate] ratio[14–16]. To verify this possibility we assessed the acute effects of metformin on fasting plasma glucose concentrations and rates of EGP at doses that are in the range used to treat patients with type 2 diabetes (20 mg kg^{-1} and 50 mg kg^{-1}) in awake rats. As early as 30 min after administration, there was a marked reduction in fasting plasma glucose concentrations (Fig. 1a) and inhibition of EGP (Fig. 1b and Extended Data Fig. 3a). We measured the [lactate]:[pyruvate] and [β-hydroxybutyrate]:[acetoacetate] ratios as surrogates of the cytosolic and mitochondrial redox states, respectively[14–16], and found that metformin elicited a significant increase in the cytosolic redox state and decrease in the mitochondrial redox state in both plasma and liver (Fig. 1c–f and Extended Data Fig. 3b, c). Consistent with these findings, metformin treatment increased the liver [GSH]:[GSSG] ratio, reflecting an increase in the cytosolic redox state (Extended Data Fig. 3d). These data support a mechanism for metformin's rapid effect on hepatic glucose production through alteration of the redox state. We also observed a trend towards increased plasma glycerol concentrations, suggesting inhibition of glycerol conversion to glucose (Fig. 1g). In contrast, acute metformin treatment did not affect cellular energy charge ([ATP]:[ADP], [ATP]:[AMP], [NADH]:[NAD^+], [NADPH]:[$NADP^+$]) (Extended Data Fig. 3e–h), or liver [cAMP] (Extended Data Fig. 3i). The lack of observed differences in [NADH]:[NAD^+] and [NADPH]:[$NADP^+$] ratios, despite significant metformin-induced alterations in the cytoplasmic and mitochondrial redox state, can most probably be attributed to the [NADH]:[NAD^+] ratio moving in opposite directions in these two compartments such that total

[1]Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [2]Department of Cellular & Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [3]Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [4]Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [5]Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523, USA. [6]Cancer Prevention Research Institute of Texas Scholar, Department of Chemistry and Biochemistry, Baylor University, Waco, Texas 76798, USA. [7]Isis Pharmaceuticals, 2855 Gazelle Court, Carlsbad, California 92010, USA. [8]University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA, 53706. [9]Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark, DK-2200.

©2014 Macmillan Publishers Limited. All rights reserved

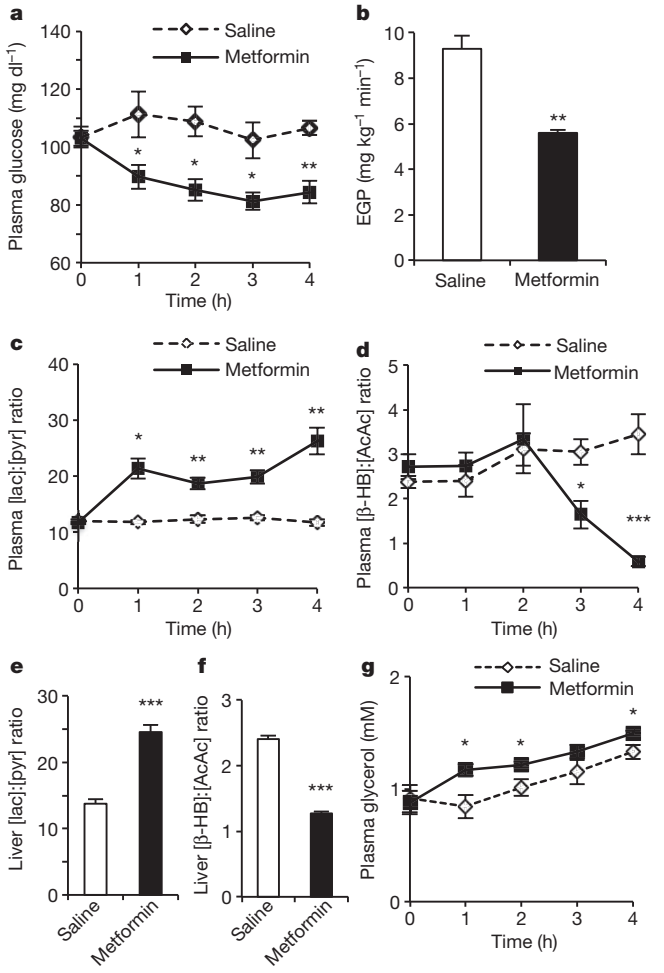


Figure 1 | Effects of acute metformin (50 mg kg^{-1}, intravenously) treatment in Sprague-Dawley rats on plasma glucose, EGP and hepatic redox state. **a**, **b**, Metformin acutely reduced (**a**) fasting plasma glucose and (**b**) EGP. These changes were associated with (**c**) increased plasma [lactate]:[pyruvate] ([lac]:[pyr]) and (**d**) decreased plasma [β-hydroxybutyrate]:[acetoacetate] ([β-HB]:[AcAc]). **e**, Increased hepatic cytosolic redox state, (**f**) decreased liver mitochondrial redox state and (**g**) a tendency to increased plasma glycerol were observed. Data set shown is representative of three experiments. Data are mean ± s.e.m. ((**a**) saline $n = 3$, metformin $n = 4$; (**b**–**d**, **g**) $n = 4$; (**e**, **f**) $n = 6$; biological replicates; experiment replicated three times). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired t-test.

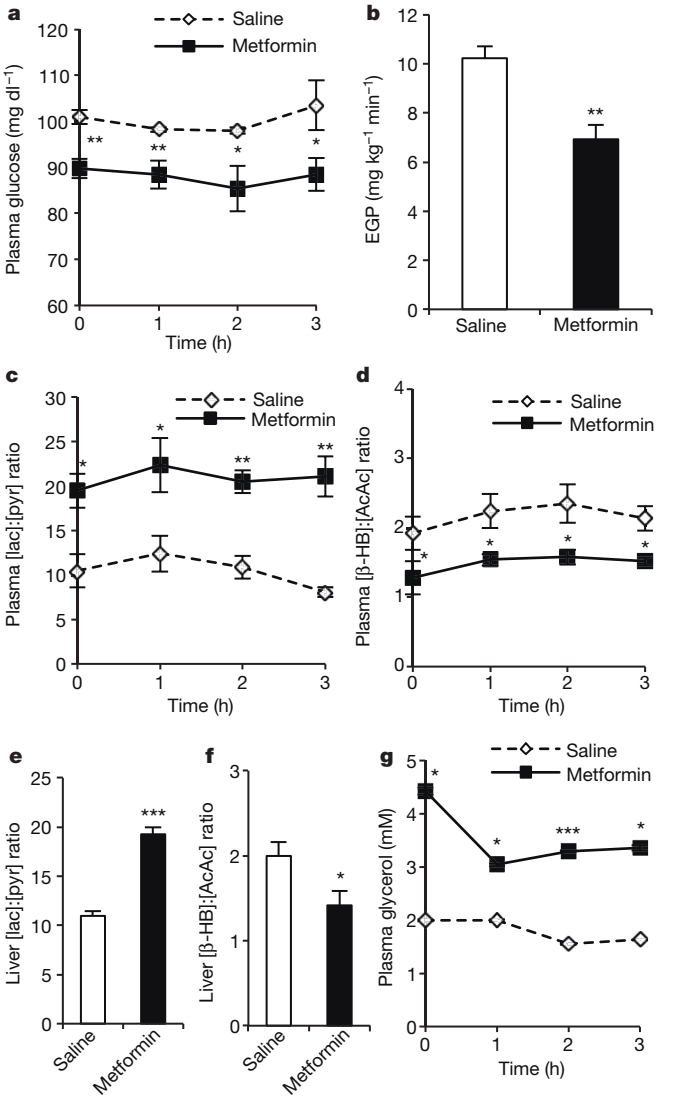


Figure 2 | Effects of chronic metformin (50 mg kg^{-1}, intraperitoneally per day × 30 days) treatment in Sprague-Dawley rats on plasma glucose, EGP and hepatic redox state. **a**, **b**, Chronic metformin treatment reduced (**a**) fasting plasma glucose and (**b**) EGP. These changes were associated with (**c**) increased plasma [lactate]:[pyruvate], and (**d**) decreased plasma [β-hydroxybutyrate]:[acetoacetate]. Chronic metformin treatment (**e**) elevated hepatic cytosolic redox state, (**f**) decreased liver mitochondrial redox state and (**g**) increased plasma glycerol. Data are mean ± s.e.m. ((**a**) $n = 5$; (**b**, **d**–**g**) saline $n = 6$, metformin $n = 4$; (**c**) $n = 4$; biological replicates). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired t-test.

[NADH]:[NAD^{+}] may remain the same. Finally, no differences in protein expression of the key gluconeogenic enzymes cytosolic phosphoenolpyruvate carboxykinase (PEPCK-C) and pyruvate carboxylase were observed (Extended Data Fig. 3j–l). There was no significant AMPK activation, as determined by relative levels of phosphorylated AMPK to total AMPK, and no change in the phosphorylation of the downstream target of AMPK, acetyl-coenzyme A carboxylase (ACC), although cyclic AMP response element binding protein (CREB) activation was observed as determined by the ratio of phosphorylated CREB to total CREB (Extended Data Fig. 3j, m–o).

Similar to acute metformin treatment, chronic metformin treatment lowered fasting plasma glucose concentrations and EGP (Fig. 2a, b). These changes were also associated with an increase in the plasma and cytosolic redox state and a decrease in the mitochondrial redox state (Fig. 2c–f and Extended Data Fig. 4a). Plasma glycerol concentrations were also higher in these animals (Fig. 2g). Similar to what was observed with acute metformin treatment, chronic metformin treatment did not alter hepatocellular energy charge (Extended Data Fig. 4b–e), and there was only a slight reduction in liver [cAMP] (Extended Data Fig. 4f). Protein expression of PEPCK-C and pyruvate carboxylase were also unchanged; however, AMPK was activated, ACC phosphorylation was increased and CREB activation was decreased (Extended Data Fig. 4g–l).

Given that acute and chronic metformin treatment increased the cytoplasmic redox state and decreased the mitochondrial redox state, we next considered that metformin may block one of the two main redox shuttle systems, the malate–aspartate and glycerophosphate shuttles[17–20]. Guanides/biguanides had no effect on total malate–aspartate shuttle activity or on malate dehydrogenase and aspartate aminotransferase activities (Extended Data Fig. 2e–g).

We next examined the effect of metformin on the glycerophosphate shuttle[21] (Extended Data Fig. 5a). Guanides/biguanides did not alter the activity of purified, recombinant cytosolic glycerophosphate dehydrogenase (cGPD) (Extended Data Fig. 5b). In contrast, all guanides/biguanides inhibited mitochondrial GPD (mGPD) activity by 30–50% in rat mitochondrial lysates (Fig. 3a). These findings are supported by our data showing

©2014 Macmillan Publishers Limited. All rights reserved

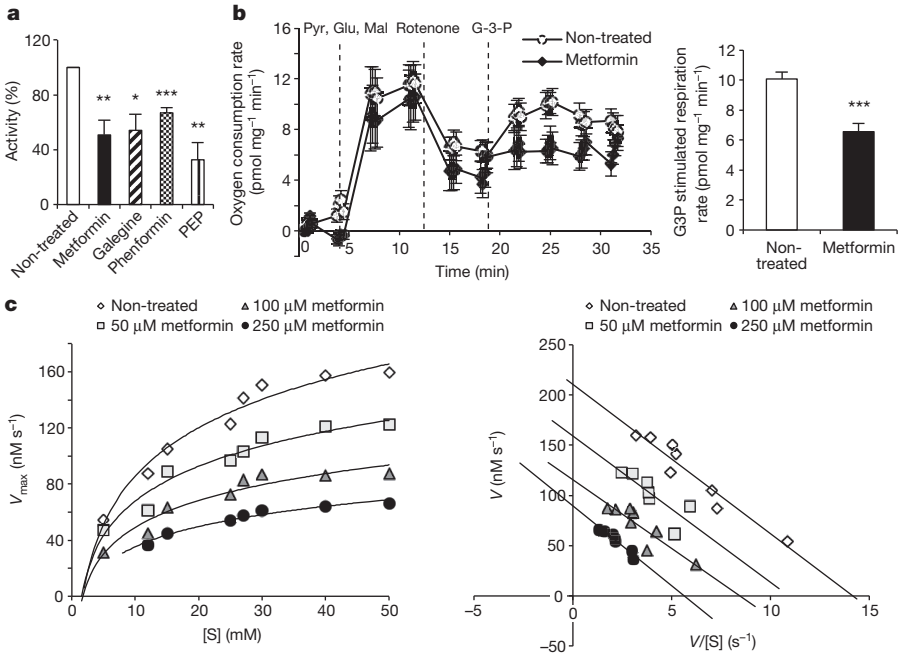


Figure 3 | Guanides/biguanides non-competitively inhibit rat and human mGPD activity. **a**, Guanides/biguanides inhibit rat mGPD activity by 30–50% in rat mitochondrial lysate. Data shown are the average of three experiments. **b**, Metformin (50 μM) inhibited mitochondrial respiration from glycerol-3-phosphate (G-3-P). Pyr, pyruvate; Glu, glutamate; Mal, malate. **c**, Metformin inhibited recombinant human mGPD non-competitively as determined by Michaelis–Menten and Eadie–Hofstee plots of the kinetic data. Data shown are representative of two experiments. Data are mean ± s.e.m. ($n = 5$ technical replicates for mGPD activity in lysate, experiment replicated three times; non-treated $n = 12$, metformin $n = 12$ technical replicates for oxygen consumption, experiment replicated twice). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by unpaired t-test.

that metformin significantly inhibited mitochondrial respiration from glycerol-3-phosphate (Fig. 3b).

Given the impact of metformin on the cytosolic/mitochondrial redox state, we next examined whether we could influence the effectiveness of metformin by changing the cytosolic redox state by altering the extracellular [lactate]:[pyruvate] ratio[22,23]. Consistent with this hypothesis we found that at lower [lactate]:[pyruvate], metformin and mGPD short interfering RNA (siRNA) treatment to knockdown expression of *mGpd* no longer significantly reduced glucose production from primary rat hepatocytes (Extended Data Fig. 6a). This finding suggests that an augmented cytosolic redox state is central to metformin-mediated reduction in hepatic gluconeogenesis.

Inhibition of mGPD would prevent glycerol from contributing to gluconeogenic flux, as glycerol must be phosphorylated by glycerol kinase to glycerol-3-phosphate and converted to dihydroyacetone phosphate by mGPD to contribute to glucose production. Furthermore, inhibition of mGPD would halt the glycerophosphate shuttle and lead to accumulation of cytosolic NADH, which is unfavourable for conversion of lactate to pyruvate by lactate dehydrogenase (LDH). Consistent with this hypothesis, glucose production in hepatocytes from both lactate and glycerol was inhibited by metformin and *mGpd* knockdown. In contrast, metformin and *mGpd* knockdown had no effect on gluconeogenesis from dihydroxyacetone, alanine and pyruvate, substrates that enter gluconeogenesis without augmenting the cytosolic [NADH]:[NAD^+] ratio (Extended Data Fig. 6b, c). We also observed an increase in glycerol-3-phosphate concentrations in primary hepatocytes after metformin treatment and *mGpd* knockdown by siRNA, and specifically in liver after acute metformin treatment *in vivo* (Extended Data Fig. 6d, e), consistent with metformin inhibition of mGPD activity. These data indicate that both acute and chronic metformin treatment inhibit mGPD, limiting lactate and glycerol contributions to hepatic gluconeogenesis.

To examine whether mGPD is the molecular target for metformin *in vivo*, we treated rats with antisense oligonucleotides (ASOs) to knock down hepatic *mGpd* and *cGpd* expression (Extended Data Fig. 7a, b). As seen with metformin treatment, mGPD ASO treatment reduced plasma glucose concentrations, increased cytosolic redox state and decreased mitochondrial redox state. In contrast, cGPD ASO treatment had only a mild effect on cytosolic and mitochondrial redox state, did not alter plasma glucose levels and did not affect plasma lactate concentrations (Fig. 4a–c and Extended Data Fig. 7c). These data are consistent with previous studies showing that knockout of the *mGpd* gene, but not *cGpd*, in mice results in a reduction in plasma glucose concentrations[24,25].

To investigate further whether mGPD is the primary target of metformin, we examined whether mGPD ASO treatment would abrogate the inhibitory effects of metformin on EGP. mGPD ASO treatment lowered plasma glucose concentrations and inhibited EGP. Moreover, mGPD ASO treatment abrogated the ability of metformin ability to decrease plasma glucose concentrations and inhibit EGP (Fig. 4d, e). mGPD-ASO-treated animals also had increased plasma [lactate]:[pyruvate] and decreased plasma [β-hydroxybutyrate]:[acetoacetate] ratios, whereas metformin had no extra effect on these parameters (Fig. 4f, g). As seen with metformin treatment, mGPD ASO increased the liver [GSH]:[GSSG] ratio, reflecting an increase in the cytosolic redox state (Extended Data Fig. 7d). In contrast the energy charge in the liver was unaltered by *mGpd* knockdown (Extended Data Fig. 7e–h), although a slight decrease in liver [cAMP] was observed (Extended Data Fig. 7i). There was no appreciable effect of *mGpd* knockdown on hepatic PEPCK-C and pyruvate carboxylase expression, although AMPK activation, increased ACC phosphorylation and suppression of CREB activation were observed (Extended Data Fig. 7j–o). Chronic metformin treatment of *mGpd* ASO knockdown rats also had no added effect on plasma glucose concentrations or EGP (Fig. 4h, i). We next examined whether metformin's effects would be abrogated in a whole-body mGPD knockout mouse and found that these mice, similar to the mGPD-ASO-treated rats, had decreased fasting plasma glucose concentrations and decreased rates of EGP compared with wild-type littermates, and that the effects of acute metformin treatment on plasma glucose concentrations and EGP were also abrogated in mGPD knockout mice (Fig. 4j, k).

We further characterized the nature of metformin's effect on mGPD kinetics, and found that metformin inhibited isolated rat mGPD activity non-competitively with an inhibition constant (K_i) of approximately 38 μM (Extended Data Fig. 5c). Metformin also inhibited purified recombinant human mGPD activity, showing non-competitive kinetics with $K_i \sim 55$ μM (Fig. 3c and Extended Data Fig. 5d), and non-competitively inhibited activity of pure, *Pediococcus* α-glycerophosphate oxidase, a bacterial isoform of mammalian mGPD (Extended Data Fig. 5e).

The glycerophosphate shuttle has an important role in the regulation of hepatic gluconeogenesis[26,27], the role of mGPD as a regulator of EGP being strongly supported by the finding that mGPD knockout mice have lowered plasma glucose concentrations[28] and are protected

©2014 Macmillan Publishers Limited. All rights reserved

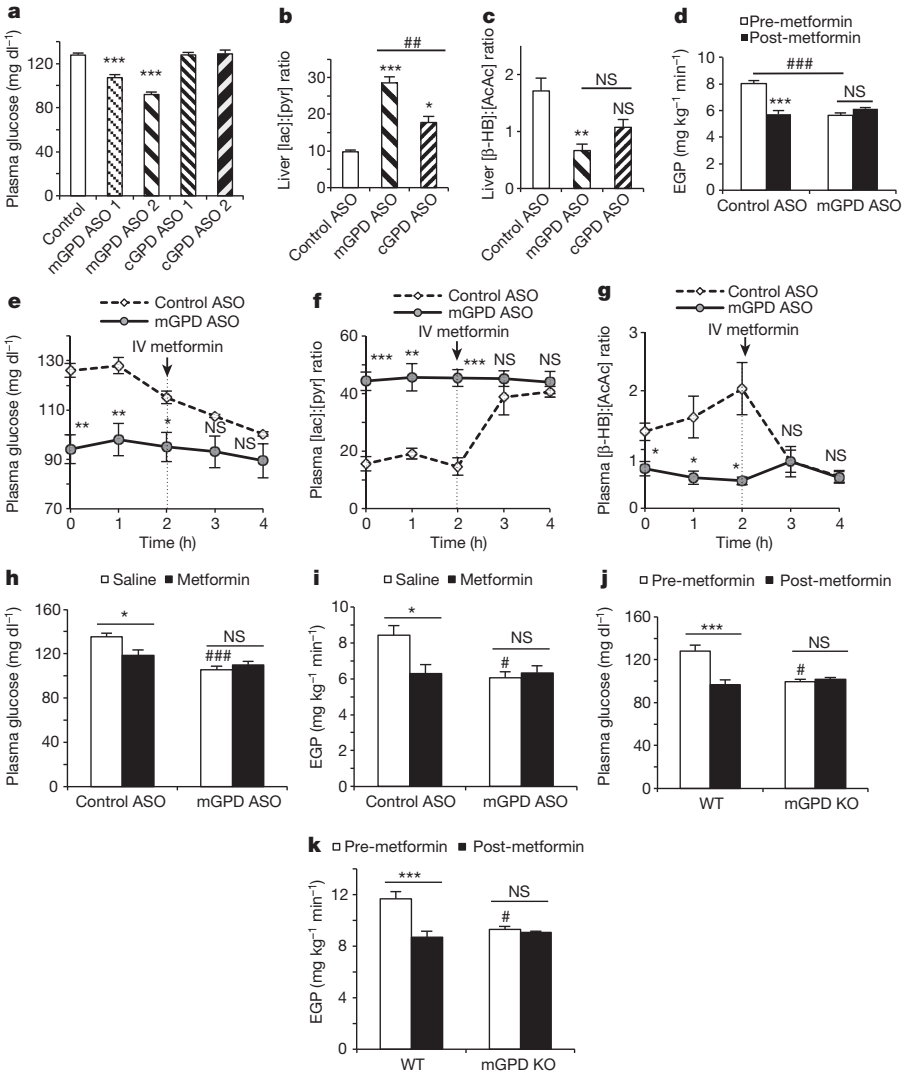


Figure 4 | Effects of mGPD and cGPD ASO treatment in Sprague-Dawley rats on metformin action. **a**, Knockdown of hepatic mGPD, but not cGPD, by two ASOs reduced fasting plasma glucose. **b**, mGPD ASO treatment increased liver [lactate]:[pyruvate], and (**c**) decreased liver [β-hydroxybutyrate]:[acetoacetate]. cGPD ASO treatment had a less marked effect on both liver redox ratios. **d**, *mGpd* knockdown decreased EGP and abrogated metformin's effect on EGP. **e**, Knockdown of hepatic mGPD decreased plasma glucose, (**f**) increased plasma [lactate]:[pyruvate], (**g**) decreased plasma [β-hydroxybutyrate]:[acetoacetate] and abrogated metformin's effect on these parameters. **h**, In mGPD-ASO-treated rats, chronic metformin treatment had no effect on plasma glucose, (**i**) or EGP. mGPD knockout mice had (**j**) decreased plasma glucose, (**k**) reduced EGP, and did not respond to acute metformin treatment. WT, wild type; KO, knockout. Data are mean ± s.e.m. ((**a**) $n = 6$; (**b**) Control $n = 6$, mGPD $n = 5$, cGPD $n = 5$; (**c**) Control $n = 6$, mGPD $n = 5$, cGPD $n = 6$; (**d**) $n = 5$; (**e**) Control $n = 4$, mGPD $n = 6$; (f,g) Control $n = 4$, mGPD $n = 5$; (h,i) $n = 7$; (**j**) WT $n = 13$, KO $n = 4$; (**k**) WT pre-metformin $n = 9$, WT metformin $n = 13$, KO $n = 4$; biological replicates). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by analysis of variance (ANOVA); $\#P < 0.05$, $\#\#P < 0.01$, $\#\#\#P < 0.001$ by ANOVA. NS, not significant.

from diet-induced hyperglycaemia[29]. We found that metformin inhibits mGPD activity *in vitro* with an approximate K_i value close to the observed range of plasma metformin levels (10–40 μM) in patients treated with a normally prescribed dose of 1 g of metformin twice a day[11]. Our measurements of metformin concentrations in the plasma and liver of rats treated acutely with 50 mg kg^{-1} metformin gave us a peak average value of 74 μM and 100 μM, respectively, 30 min after administration (Extended Data Fig. 9). These pharmacokinetic data are consistent with the acute metformin-mediated inhibition of hepatic gluconeogenesis within a similar time frame and are consistent with the need to provide dosing of metformin twice a day to patients with type 2 diabetes for optimal plasma glucose lowering. Previous studies assessing metformin action *in vivo* have typically used metformin doses ranging from 250 mg kg^{-1} to 500 mg kg^{-1}, which are supra-pharmacological and result in plasma metformin concentrations greater than 1 mM (Extended Data Fig. 9). In this regard, although metformin inhibited mitochondrial respiration from glycerol-3-phosphate at micromolar concentrations, complex I respiration was affected only at metformin concentrations greater than 1 mM, which are not clinically relevant (Extended Data Fig. 2h).

The kinetics of metformin inhibition of mGPD was found to be non-competitive, which raises the question of how this might be achieved. Computational analyses of the structure of the mGPD isoform, *Streptococcus* sp. α-glycerophosphate oxidase[30], mutated in the flavin adenine dinucleotide (FAD) binding pocket to the sequence of the human mGPD, found that metformin shows favourable binding to the FAD-containing pocket (Extended Data Fig. 8a–c). Based on these predictions, FAD, which is non-covalently associated with the enzyme, may dynamically slide in and out of the wider entrance of the tunnel-shaped pocket, and metformin may enter from the smaller end, preventing complete re-entry of FAD and inhibiting electron transfer. Although speculative, this model could explain how metformin could displace the flavin ring and inhibit enzymatic activity without necessarily invoking FAD turnover.

In summary, we have shown that metformin treatment of rats at doses that achieve comparable plasma metformin concentrations observed in patients with metformin-treated type 2 diabetes inhibits mGPD non-competitively and modulates cytosolic and mitochondrial redox state, inducing an effective reduction in EGP. These results identify mGPD as one of the primary molecular targets by which guanides/biguanides inhibit hepatic gluconeogenesis, and provide a new therapeutic target for type 2 diabetes.

METHODS SUMMARY

Hepatocyte studies. Primary hepatocytes were isolated from 24-h-fasted Sprague-Dawley rats and treated with or without 100 μM metformin in DMEM supplemented with 10 mM substrate (glycerol at 100 μM). Glucose was measured using Genzyme Glucose-SL reagent.

Animal studies. Studies were performed in 24-h-fasted, awake, unrestrained male Sprague-Dawley rats aged 9–12 weeks. Animals were randomly allocated to treatment groups before collection of any data and the studies were performed unblinded. Acute metformin was administered intravenously and chronic metformin was administered intraperitoneally at 50 mg kg^{-1} daily over 30 days. Continuous 1 mg kg^{-1} min^{-1} arterial infusion of [6,6-$^{2}H_2$]glucose tracer was administered to measure EGP. Plasma

©2014 Macmillan Publishers Limited. All rights reserved

glucose was measured on a YSI 2700 Biochemistry Analyzer. All other metabolites were measured using enzymatic methods. Whole-body mGPD knockout mouse embryos were a gift from M. J. MacDonald. Overnight-fasted mice were infused with [3-^{3}H]glucose at 0.05 μCi min^{-1} for 120 min to measure basal glucose turnover. Mice were restrained during the studies, and given acute metformin (50 mg kg^{-1}, intravenously) or saline. All animal studies were conducted with previous approval from Yale University Institutional Animal Care and Use Committee.

ASO treatment studies. ASO treatment was administered at 37.5 mg kg^{-1} twice weekly for 4 weeks by intraperitoneal injection. ASO treatment knocked down mGPD messenger RNA (mRNA) and protein only in the liver (Extended Data Fig. 10).

Enzyme assays. mGPD activity in rat liver mitochondrial lysate was measured using cytochrome *c* as the electron acceptor and measuring gain of absorbance at 550 nm. Activities of rat mGPD isolated from rat mitochondrial lysate using anti-GPD2 antibody K-14, and human mGPD overexpressed in HEK 293 cells from an Origene vector, were measured using 2,6-dichloroindophenol (DCIP) and measuring loss of absorbance at 600 nm. Kinetic measurements were made using a Flexstation 3 Benchtop Multi-Mode Microplate Reader (Molecular Devices).

Oxygen consumption. Studies were performed on a Seahorse Bioscience XF24 Analyzer.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 May 2013; accepted 21 March 2014.
Published online 21 May 2014.

1. Hundal, R. S. *et al.* Mechanism by which metformin reduces glucose production in type 2 diabetes. *Diabetes* **49,** 2063–2069 (2000).
2. Inzucchi, S, *et al.* Efficacy and metabolic effects of metformin and troglitazone in type II diabetes mellitus. *N. Engl. J. Med.* **338,** 867–872 (1998).
3. El-Mir, M.-Y. *et al.* Dimethylbiguanide inhibits cell respiration via an indirect effect targeted on the respiratory chain complex I. *J. Biol. Chem.* **275,** 223–228 (2000).
4. Owen, M. R., Doran, E. & Halestrap, A. P. Evidence that metformin exerts its anti-diabetic effects through inhibition of complex 1 of the mitochondrial respiratory chain. *Biochem. J.* **348,** 607–614 (2000).
5. Zhou, G. *et al.* Role of AMP-activated protein kinase in mechanism of metformin action. *J. Clin. Invest.* **108,** 1167–1174 (2001).
6. Shaw, R. J. *et al.* The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin. *Science* **310,** 1642–1646 (2005).
7. He, L. *et al.* Metformin and insulin suppress hepatic gluconeogenesis through phosphorylation of CREB binding protein. *Cell* **137,** 635–646 (2009).
8. Cool, B. *et al.* Identification and characterization of a small molecule AMPK activator that treats key components of type 2 diabetes and the metabolic syndrome. *Cell Metab.* **3,** 403–416 (2006).
9. Savage, D. B. *et al.* Reversal of diet-induced hepatic steatosis and hepatic insulin resistance by antisense oligonucleotide inhibitors of acetyl-CoA carboxylases 1 and 2. *J. Clin. Invest.* **116,** 817–824 (2006).
10. Fullerton, M. D. *et al.* Single phosphorylation sites in Acc1 and Acc2 regulate lipid homeostasis and the insulin-sensitizing effects of metformin. *Nature Med.* **19,** 1649–1654 (2013).
11. Foretz, M. *et al.* Metformin inhibits hepatic gluconeogenesis in mice independently of the LKB1/AMPK pathway via a decrease in hepatic energy state. *J. Clin. Invest.* **120,** 2355–2369 (2010).
12. Miller, R. A. *et al.* Biguanides suppress hepatic glucagon signalling by decreasing production of cyclic AMP. *Nature* **494,** 256–260 (2013).
13. Hawley, S. A., Gadalla, A. E., Olsen, G. S. & Hardie, D. G. The antidiabetic drug metformin activates the AMP-activated protein kinase cascade via an adenine nucleotide-independent mechanism. *Diabetes* **51,** 2420–2425 (2002).
14. Yang, L. *et al.* Metabolomic and mass isotopomer analysis of liver gluconeogenesis and citric acid cycle. I. Interrelation between gluconeogenesis and cataplerosis; formation of methoxamates from aminooxyacetate and ketoacids. *J. Biol. Chem.* **283,** 21978–21987 (2008).
15. Krebs, H. A. & Gascoyne, T. The redox state of the nicotinamide-adenine dinucleotides in rat liver homogenates. *Biochem. J.* **108,** 513–520 (1968).
16. Williamson, D. H., Lund, P. & Krebs, H. A. The redox state of free nicotinamide-adenine dinucleotide in the cytoplasm and mitochondria of rat liver. *Biochem. J.* **103,** 514–527 (1967).
17. Bremer, J. & Davis, E. J. Studies on the active transfer of reducing equivalents into mitochondria via the malate-aspartate shuttle. *Biochim. Biophys. Acta* **376,** 387–397 (1975).
18. Cederbaum, A. I., Lieber, C. S., Beattie, D. S. & Rubin, E. Characterization of shuttle mechanisms for the transport of reducing equivalents into mitochondria. *Arch. Biochem. Biophys.* **158,** 763–781 (1973).
19. Garrib A. & McMurray, W. C. Purification and characterization of glycerol-3-phosphate dehydrogenase (flavin-linked) from rat liver mitochondria. *J. Biol. Chem.* **261,** 8042–8048 (1986).
20. Cole, E. S., Lepp, C. A., Holohan, P. D. & Fondy, T. P. Isolation and characterization of flavin-linked glycerol-3-phosphate dehydrogenase from rabbit skeletal muscle mitochondria and comparison with the enzyme from rabbit brain. *J. Biol. Chem.* **253,** 7952–7959 (1978).
21. White, H. B. III & Kaplan, N. O. Purification and properties of two types of diphosphopyridine nucleotide-linked glycerol 3-phosphate dehydrogenases from chicken breast muscle and chicken liver. *J. Biol. Chem.* **244,** 6031–6039 (1969).
22. Sistare, F. D. & Haynes R. C. Jr. The interaction between the cytosolic pyridine nucleotide redox potential and gluconeogenesis from lactate/pyruvate in isolated rat hepatocytes. *J. Biol. Chem.* **260,** 12748–12753 (1985).
23. Sugano, T. *et al.* Intracellular redox state and stimulation of gluconeogenesis by glucagon and norepinephrine in the perfused rat liver. *J. Biochem.* **87,** 153–166 (1980).
24. MacDonald, M. J. & Marshall, L. K. Mouse lacking NAD+-linked glycerol phosphate dehydrogenase has normal pancreatic beta cell function but abnormal metabolite pattern in skeletal muscle. *Arch. Biochem. Biophys.* **384,** 143–153 (2000).
25. Prochazka, M., Kozak, U. C. & Kozak, L. P. A glycerol-3-phosphate dehydrogenase null mutant in BALB/cHeA mice. *J. Biol. Chem.* **264,** 4679–4683 (1989).
26. Harding, J. W. Jr, Pyeritz, E. A., Copeland, E. S. & White, H. B. III. Role of glycerol 3-phosphate dehydrogenase in glyceride metabolism. Effect of diet on enzyme activities in chicken liver. *Biochem. J.* **146,** 223–229 (1975).
27. Harding, J. W. Jr, Pyeritz, E. A., Morris, H. P. & White, H. B. III. Proportional activities of glycerol kinase and glycerol 3-phosphate dehydrogenase in rat hepatomas. *Biochem. J.* **148,** 545–550 (1975).
28. Brown, L. J. *et al.* Normal thyroid thermogenesis but reduced viability and adiposity in mice lacking the mitochondrial glycerol phosphate dehydrogenase. *J. Biol. Chem.* **277,** 32892–32898 (2002).
29. Barberà, A. *et al.* A high carbohydrate diet does not induce hyperglycaemia in a mitochondrial glycerol-3-phosphate dehydrogenase-deficient mouse. *Diabetologia* **46,** 1394–1401 (2003).
30. Colussi, T. *et al.* Structure of α-glycerophosphate oxidase from *Streptococcus* sp.: a template for the mitochondrial α-glycerophosphate dehydrogenase. *Biochemistry* **47,** 965–977 (2008).

Acknowledgements We thank J. Dong, M. Kahn, Y. Kosover and J. J. Hsiao for their technical support, and S. Singh and D. Gregg for discussions. This publication was supported by grants from the National Institutes of Health: R24 DK-085638, R01 DK-40936, P30 DK-45735, P30 DK-034989, U24 DK-059635, R01 DK-28348, K01 DK-099402, R01 DK-092606; grants from the American Diabetes Association: 7-12-BS-092, 1-14-Merck-10; a VA Merit Award (I01-BX000901); and by the Novo Nordisk Foundation Center for Basic Metabolic Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the National Center for Research Resources or National Institutes of Health.

Author Contributions A.K.M., D.M.E., V.T.S., R.G.K., J.-P.C., M.J.J., H.-Y.L. and G.I.S. designed the experimental protocols. A.K.M., D.M.E., Y.R., J.H., X-M.Z., G.W.C., R.G.K., J.-P.C., M.J.J. and H.-Y.L. performed the studies. A.K.M., D.M.E., X-M.Z., D.B., R.A.A., G.W.C., J.-P.C., M.J.J. and H.-Y.L. analysed the data. B.J.P., J.L.W., S.B. and M.J.M. supplied reagents. A.K.M., D.M.E., Y.R., G.W.C., V.T.S., R.G.K., R.A.A., M.J.M. and G.I.S. contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.I.S. (gerald.shulman@yale.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13267

Cyclin D1–Cdk4 controls glucose metabolism independently of cell cycle progression

Yoonjin Lee[1,2,3], John E. Dominy[1,2], Yoon Jong Choi[1,4], Michael Jurczak[5], Nicola Tolliday[6], Joao Paulo Camporez[5], Helen Chim[1,2], Ji-Hong Lim[1,2], Hai-Bin Ruan[5], Xiaoyong Yang[5], Francisca Vazquez[1,2], Piotr Sicinski[1,4], Gerald I. Shulman[5] & Pere Puigserver[1,2]

Insulin constitutes a principal evolutionarily conserved hormonal axis for maintaining glucose homeostasis[1–3]; dysregulation of this axis causes diabetes[2,4]. PGC-1α (peroxisome-proliferator-activated receptor-γ coactivator-1α) links insulin signalling to the expression of glucose and lipid metabolic genes[5–7]. The histone acetyltransferase GCN5 (general control non-repressed protein 5) acetylates PGC-1α and suppresses its transcriptional activity, whereas sirtuin 1 deacetylates and activates PGC-1α[8,9]. Although insulin is a mitogenic signal in proliferative cells[10,11], whether components of the cell cycle machinery contribute to its metabolic action is poorly understood. Here we report that in mice insulin activates cyclin D1–cyclin-dependent kinase 4 (Cdk4), which, in turn, increases GCN5 acetyltransferase activity and suppresses hepatic glucose production independently of cell cycle progression. Through a cell-based high-throughput chemical screen, we identify a Cdk4 inhibitor that potently decreases PGC-1α acetylation. Insulin/GSK-3β (glycogen synthase kinase 3-beta) signalling induces cyclin D1 protein stability by sequestering cyclin D1 in the nucleus. In parallel, dietary amino acids increase hepatic cyclin D1 messenger RNA transcripts. Activated cyclin D1–Cdk4 kinase phosphorylates and activates GCN5, which then acetylates and inhibits PGC-1α activity on gluconeogenic genes. Loss of hepatic cyclin D1 results in increased gluconeogenesis and hyperglycaemia. In diabetic models, cyclin D1–Cdk4 is chronically elevated and refractory to fasting/feeding transitions; nevertheless further activation of this kinase normalizes glycaemia. Our findings show that insulin uses components of the cell cycle machinery in post-mitotic cells to control glucose homeostasis independently of cell division.

To discover new factors that can regulate PGC-1α activity through its acetylation status, a high-throughput enzyme-linked immunoassay was designed to monitor specifically and quantitatively the amount of PGC-1α acetylation in U-2OS cells (Extended Data Fig. 1a). A library of 1,600 compounds, including bioactive and natural compounds, was screened (Fig. 1a). Interestingly, the compound with the highest *z* score for PGC-1α deacetylation was fascaplysin, a known Cdk4 inhibitor[12] (Extended Data Fig. 1b). Cdk4 regulates G1 to S phase transition and its kinase activity is dependent on its binding to one of the three D-type cyclins including cyclin D1 (ref. 13). We therefore investigated the effect of this cell cycle complex on PGC-1α acetylation and function, in connection with nutrient and insulin metabolic actions.

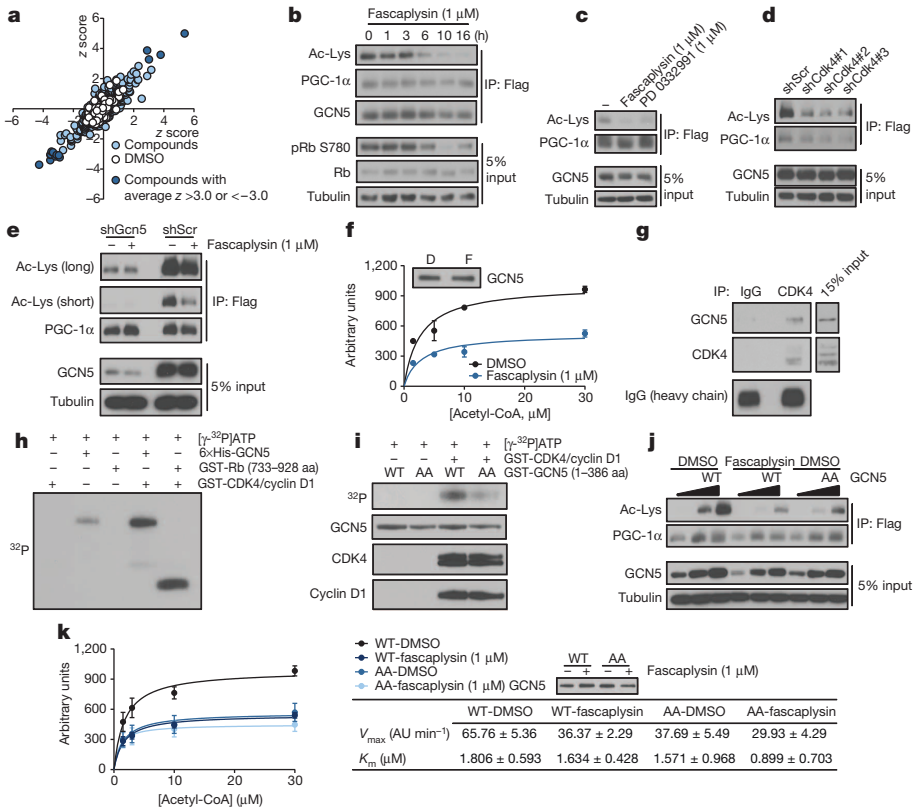
First, we calculated a half-maximum inhibitory concentration (IC_{50}) of 0.7 μM for fascaplysin-induced PGC-1α deaceylation, which is similar to its IC_{50} for Cdk4 inhibition (Extended Data Fig. 2a). Fascaplysin-induced PGC-1α deacetylation overlapped with retinoblastoma protein (Rb) dephosphorylation, a well-characterized Cdk4 substrate[14] (Fig. 1b). PD 0332991, the most specific Cdk4 inhibitor available[15], led to a similar decrease of PGC-1α acetylation (Fig. 1c and Extended Data Fig. 2b). Furthermore, Cdk4 depletion through transient short hairpin RNA (shRNA) transfection had the same effect as chemical inhibitors, confirming that Cdk4 activity controls PGC-1α acetylation levels (Fig. 1d and Extended Data Fig. 2c).

Because Cdk4 inhibitor-induced PGC-1α deacetylation was not affected when sirtuin 1 (SIRT1) or HDAC class I/II were inhibited (Extended Data Fig. 2d), we tested whether cyclin D1–Cdk4 regulates PGC-1α acetylation through GCN5, the principal PGC-1α acetyltransferase. Indeed, knockdown of *Gcn5* significantly blunted fascaplysin-induced PGC-1α deacetylation (Fig. 1e). In contrast, PCAF-mediated acetylation was not affected by fascaplysin, further suggesting that Cdk4 inhibition modulates PGC-1α acetylation through GCN5 (Extended Data Fig. 2e). *In vitro* catalytic activity of GCN5 immunoprecipitated from cells treated with fascaplysin was reduced by 50% relative to vehicle control (Fig. 1f). We observed physical interaction between ectopically expressed or endogenous Cdk4 and GCN5, suggesting that Cdk4 could regulate GCN5 activity by direct phosphorylation (Fig. 1g and Extended Data Fig. 2f). Cyclin D1–Cdk4 kinase directly phosphorylated GCN5 *in vitro* and its phosphorylation was inhibited by fascaplysin (Fig. 1h and Extended Data Fig. 2g). Systematic mutagenesis showed two phosphorylation sites, T272 and S372, located in the conserved PCAF domain of GCN5. Alanine substitutions of these two sites ('GCN5 AA') ablated GCN5 phosphorylation by cyclin D1–Cdk4 *in vitro*, and reduced PGC-1α acetylation (Fig. 1i, 1j and Extended Data Fig. 2h, i). Compared with GCN5 wild type, *in vitro* catalytic activity of GCN5 AA was decreased, but remained insensitive to fascaplysin (Fig. 1k). Cdk4 phosphorylation on GCN5 augmented acetyltransferase catalytic activity by increasing maximal velocity (V_{max}), whereas the Michaelis–Menten constant (K_m) for acetyl-CoA binding was unaffected (Fig. 1k). Because GCN5 functions as a complex with subunits important for its activity[16], its phosphorylation by Cdk4 could modulate interaction between GCN5 and subunits. We found one subunit, PAF65-β, interacting less with GCN5 AA compared with GCN5 wild type, when tested with modestly overexpressed GCN5, where PGC-1α acetylation was not saturated (Extended Data Fig. 2j, k). Taken together, these results indicate that cyclin D1–Cdk4 regulates PGC-1α acetylation through the direct phosphorylation and activation of GCN5 acetyltransferase activity.

Because PGC-1α acetylation is tightly linked with its co-transcriptional activity[8,9,17], we investigated whether cyclin D1–Cdk4 could modulate the gluconeogenic target genes of PGC-1α. To study cyclin D1–Cdk4 kinase effects in a PGC-1α mediated manner, we induced PGC-1α in primary hepatocytes by adenoviral overexpression or forskolin addition (Extended Data Fig. 3a). In primary hepatocytes, chemical inhibition of Cdk4 significantly increased gluconeogenic genes and glucose production, consistent with reduced PGC-1α acetylation (Fig. 2a–c and Extended Data Fig. 3b). Cdk4 depletion produced similar effects (Fig. 2d–f and Extended Data Fig. 3c). Conversely, induction of CKD4 activity by cyclin D1 wild-type overexpression or cyclin D1 T286A mutant, a hyperactive mutant allele spared from proteasomal degradation[18], but not cyclin D1 K112E (a mutant that cannot activate Cdk4 kinase activity[19]), suppressed

[1]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [2]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [5]Yale's Mouse Metabolic Phenotyping Center and Departments of Internal Medicine and Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA. [6]Chemical Biology Platform, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, Massachusetts 02141, USA.

©2014 Macmillan Publishers Limited. All rights reserved



	WT-DMSO	WT-fascaplysin	AA-DMSO	AA-fascaplysin
V_{max} (AU min^{-1})	65.76 ± 5.36	36.37 ± 2.29	37.69 ± 5.49	29.93 ± 4.29
K_m (μM)	1.806 ± 0.593	1.634 ± 0.428	1.571 ± 0.968	0.899 ± 0.703

Figure 1 | Cyclin D1–Cdk4 modulates PGC-1α acetylation through GCN5. **a**, Scatter plot of chemicals plotted with first test *z* scores on the *x* axis and repeated test scores on the *y* axis. DMSO, dimethylsulphoxide. **b**, Fascaplysin reduces PGC-1α acetylation and Rb phoshorylation. **c**, Fascaplysin and PD 0332991 treatments decrease PGC-1α acetylation. **d**, *Cdk4* knockdown causes PGC-1α deacetylation. **e**, *Gcn5* knockdown blunts fascaplysin-mediated PGC-1α deacetylation. **f**, GCN5 acetyltransferase activity is reduced upon fascaplysin treatment ($n = 2$, technical replicates, mean ± s.e.m., *y* axis represents florescence generated for 15 min from the assay). D, DMSO; F, fascaplysin. **g**, Endogenous GCN5 and Cdk4 interact. **h**, Cyclin D1–Cdk4 kinase phosphorylates GCN5 *in vitro*. aa, amino acids. **i**, GCN5 T272A/S372A (AA) phosphorylation by cyclin D1–Cdk4 kinase is diminished compared with GCN5 wild type (WT). **j**, GCN5 T272A/S372A shows decreased acetyltransferase capacity. **k**, GCN5 T272A/S372A has decreased acetyltransferase activity and insensitivity to fascaplysin treatment. Kinetic constants were calculated by the Michaelis–Menten equation ($n = 4$, two biological replicates assayed in duplicate; AU/min, arbitrary units per minute, mean ± s.e.m., *y* axis represents florescence generated for 15 min from the assay). U-2OS cells were used for these experiments.

gluconeogenic genes and glucose production, corresponding with PGC-1α acetylation changes (Fig. 2g–i and Extended Data Fig. 3d).

Because cyclin D1–Cdk4 can phosphorylate other regulatory components of gluconeogenic pathway[20] (Extended Data Fig. 3e), we examined whether cyclin D1–Cdk4 kinase modulates glucose production through PGC-1α and GCN5. As expected, the induction of gluconeogenic genes by Cdk4 inhibition was completely blunted when PGC-1α was stably knocked down in liver cells (Extended Data Fig. 3f). Also, *Cdk4* knockdown did not increase glucose production and gluconeogenic gene expression in GCN5 depleted hepatocytes (Fig. 2j and Extended Data Fig. 3g). GCN5 AA had a marginal ability to suppress hepatic gluconeogenesis compared with GCN5 wild type, and showed no significant changes when combined with Cdk4 inhibition (Fig. 2k and Extended Data Fig. 3h). Together, these results strongly suggest that the effects of cyclin D1–Cdk4 on gluconeogenesis are mediated by the GCN5-PGC-1α complex.

We examined whether liver gluconeogenesis was altered when cyclin D1–Cdk4 kinase activity was manipulated in whole animals. When PD 0332991 was administered to re-fed mice, it increased *Pck1* expression and glycaemia without changing circulating insulin levels (Fig. 2l, m and Extended Data Fig. 3i, j). In contrast, hepatic cyclin D1 T286A expression significantly repressed fasting gluconeogenic genes, glycaemia and hepatic capacity to produce glucose as assessed by a pyruvate tolerance test (Fig. 2n–p and Extended Data Fig. 3k). We confirmed that PD 03322991 or cyclin D1 T286A adenoviral administration produced those physiological changes without causing liver toxicity (Extended Data Fig. 4a, b). Collectively, these data indicate that changes in hepatic cyclin D1–Cdk4 activity are sufficient to control hepatic gluconeogenesis and whole-body glucose homeostasis.

Next, we investigated whether cyclin D1–Cdk4 complex is regulated under nutritional and hormonal changes. In spite of parenchymal hepatocytes being post-mitotic, hepatic protein and messenger RNA (mRNA) expression of cyclin D1 were significantly increased upon 4 h re-feeding, which was not a strain-specific phenomenon (Fig. 3a, b and Extended Data Fig. 5a, b). This induction of cyclin D1 corresponded with elevated cyclin D1–Cdk4 activity, as observed by increases of Rb, GCN5 phosphorylation and elevated kinase activity of immunoprecipitated Cdk4 upon re-feeding (Fig. 3b, c and Extended Data Fig. 5c). Interestingly, cyclin D1 expression was regulated upon re-feeding only in liver, and Cdk4 inhibitor administration altered PGC-1α target genes only in liver and white adipose tissue (Extended Data Fig. 5d, e). Although Cdk4 regulates glucose metabolism in other tissues than liver, it is unknown whether this regulation is dependent on PGC-1α[21].

Cyclin D1–Cdk4 activity increased upon re-feeding was uncoupled from cell cycle progression. Known cell division markers did not change upon re-feeding (Extended Data Fig. 5f). Consistently, we found no difference in hepatocyte ploidy profile, Ki-67 expression levels and 5-bromodeoxyuridine (BrdU) incorporation between the livers of fasted and re-fed mice (Fig. 3d, e and Extended Data Fig. 5g). Alteration of cyclin D1–Cdk4 activity in liver by hepatic overexpression of cyclin D1 T286A or by a pharmacological inhibition caused no changes in cell cycle progression during fasting and re-feeding, as analysed by hepatic ploidy and Ki-67 expression (Extended Data Fig. 5h–j).

GSK-3β, active during fasting and inactive during re-feeding owing to its phosphorylation by insulin-AKT signalling, negatively regulates cyclin D1 through phosphorylation of T286, causing nuclear exclusion of cyclin D1 and proteasomal degradation[18,22]. We considered that insulin stimulation and subsequent inactivation of GSK-3β allow cyclin D1 to form an active complex with Cdk4, promoting PGC-1α acetylation. Insulin or two GSK-3β inhibitors increased nuclear cyclin D1 protein when tested in an overexpressed or endogenous manner (Fig. 3f and Extended Data Fig. 5k). This increase correlated with elevated PGC-1α acetylation and a significant reduction of gluconeogenic genes (Fig. 3g and Extended Data Fig. 5l). In accordance, insulin increased Cdk4-mediated GCN5 phosphorylation (Fig. 3h). Interestingly, we found that amino acids, but not insulin, increased cyclin D1 mRNA expression in primary hepatocytes (Extended Data Fig. 5m–o). To test if amino acids could be dietary components promoting cyclin D1 mRNA in liver, mice were fed with empty calorie, glucose only or glucose and amino-acid diets for 4 h. A significant cyclin D1 mRNA induction was observed only

©2014 Macmillan Publishers Limited. All rights reserved

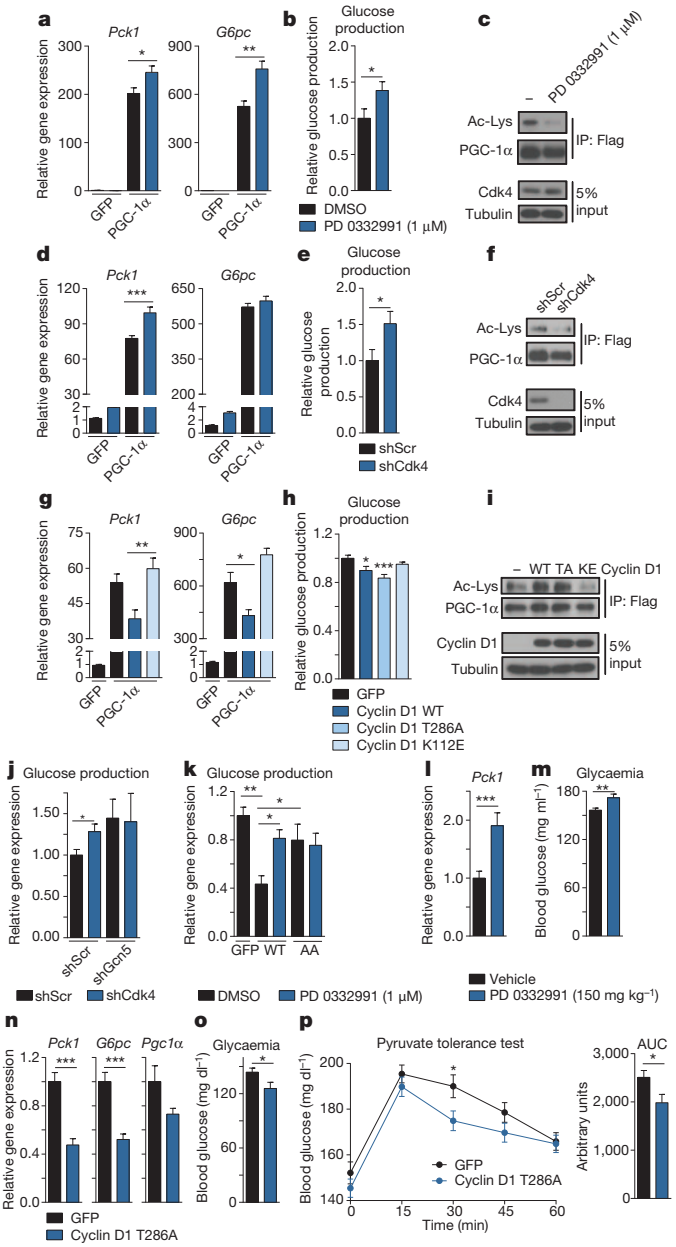


Figure 2 | Cyclin D1–Cdk4 regulates gluconeogenesis in primary hepatocytes and whole animals. **a**, **b**, PD 0332991 increases gluconeogenic gene expression and glucose production (**a**: one-way analysis of variance (ANOVA) with Tukey post-test, $n = 3$, **b**: two-tailed unpaired *t*-test, $n = 8$). **c**, PGC-1α acetylation is decreased by PD 0332991 treatment. **d**, **e**, *Cdk4* knockdown increases gluconeogenic gene expression and glucose production (**d**: one-way ANOVA with Tukey post-test, $n = 6$, **e**: two-tailed unpaired *t*-test, $n = 6$). **f**, PGC-1α acetylation is decreased upon Cdk4 deletion. **g**, **h**, Cyclin D1 wild-type decreases gluconeogenic gene expression, and cyclin D1 wild-type and cyclin D1 T286A, but not cyclin D1 K112E, repress glucose production (**g**: one-way ANOVA with Tukey post-test, $n = 3$/green fluorescent protein (GFP), PGC-1α, $n = 6$/cyclin D1 WT and KE, **h**: one-way ANOVA with Dunnett post-test, $n = 8$). **i**, Overexpression of cyclin D1 wild-type (WT) and T286A (TA), but not cyclin D1 K112E (KE), induces PGC-1α acetylation. **j**, *Gcn5* knockdown blunts the increase of glucose production by *Cdk4* knockdown (two-tailed unpaired *t*-test, $n = 8$). **k**, PD 0332991 increases glucose production with GCN5 wild type (WT), but not with GCN5 T272A/S372A (AA) (one-way ANOVA with Newman–Keuls post-test, $n = 8$). **l**, **m**, PD 0332991 administration increases *Pck1* gene expression and glycaemia in mice (two-tailed unpaired *t*-test, $n = 18$/vehicle, $n = 17$/PD 0332991). **n**, **o**, Gluconeogenic gene expression and glycaemia are reduced by cyclin D1 T286A overexpression in liver (two-tailed unpaired *t*-test, $n = 10$/GFP, $n = 9$/D1 T286A). **p**, Cyclin D1 T286A overexpression in liver decreases hepatic glucose production capacity assessed by pyruvate tolerance test (two-tailed unpaired *t*-test, $n = 20$/GFP, $n = 24$/D1 T286A; AUC, area under curve). All cells were infected with PGC-1α adenovirus for glucose production assays. Statistical significance is represented by asterisks corresponding to $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$. Data are shown as mean ± s.e.m.

when amino acids were added to the diet (Fig. 3i). These results indicate that amino acids contribute to the fed response along with insulin signalling to enhance cyclin D1 levels during re-feeding.

To test if cyclin D1–Cdk4 complex constitutes an important axis of the nutrient and insulin signalling to regulate glucose metabolism, we generated liver-specific cyclin D1 knockout ('D1 LKO') mice by crossing mice expressing a floxed cyclin D1 allele[23] with albumin-Cre-expressing mice[24]. Cyclin D1 deletion in liver did not interfere with liver development and caused no compensatory increases on other cyclins (Extended Data Fig. 5p, q). D1 LKO mice had similar gluconeogenic gene expression and glycaemia during fasting compared with wild-type mice, consistent with Cdk4 inactivation due to the absence of cyclin D1 (Fig. 3j, k). However, in the re-fed state, D1 LKO mice showed a significant increase in gluconeogenic genes and glycaemia compared with wild-type mice whereas induction of Rb phosphorylation and PGC-1α acetylation upon re-feeding were lacking (Fig. 3l, m and Extended Data Fig. 5r, s). Cdk4 inhibitor failed to increase gluconeogenic genes and glycaemia further in D1 LKO mice, suggesting cyclin D1 in liver mediates metabolic effects caused by Cdk4 inhibition (Extended Data Fig. 5t, u). Primary hepatocytes isolated from D1 LKO mice had increased gluconeogenic gene expression and higher glucose production than wild-type hepatocytes (Extended Data Fig. 5v, w). Consistent with elevated re-feeding blood glucose levels, D1 LKO mice showed moderate, yet significant, insulin and glucose intolerance, further indicating that cyclin D1–Cdk4 mediates, at least in part, insulin action in liver (Fig. 3n, o). When taken in aggregate, our findings indicate that insulin suppresses hepatic gluconeogenesis during re-feeding, in part by activating cyclin D1–Cdk4 complex to suppress PGC-1α activity.

Lastly, we investigated whether cyclin D1–Cdk4 activity was altered in diabetic mice. In the livers of fasted *Lepr*$^{db/db}$ mice, basal phosphorylations of AKT and its downstream targets were elevated, consistent with the hyperinsulinaemia of *db/db* mice (Fig. 4a). Accordingly, *db/db* mice had elevated fasted levels of cyclin D1 and phosphorylation of Rb, comparable to re-fed *db*/+ mice. In *db/db* mice, gluconeogenic genes and insulin signalling were refractory to fasting/re-feeding transitions, indicating an uncoupling of insulin signalling and its repressive effects on gluconeogenic programming (Fig. 4a and Extended Data Fig. 6a). We observed similar chronic elevations of cyclin D1 and insulin signalling impairment in mice fed a high-fat diet (Extended Data Fig. 6b). In accordance with chronic elevation of cyclin D1 in diabetic mice, phosphorylation of GCN5 remained higher, refractory to fasting/re-feeding transitions in diabetic compared with control mice (Extended Data Fig. 6c, d). These results suggest that in insulin-resistant mice compensatory hyperinsulinaemia maintains a chronically active cyclin D1–Cdk4 complex that is not sufficient to counteract increased gluconeogenesis. Hyperinsulinaemia might explain other studies showing different activities of insulin signalling in fasted diabetic animals[25,26].

We examined whether hyperactivation of cyclin D1–Cdk4 complex in *db/db* mice could alleviate hyperglycaemia. Hepatic overexpression of cyclin D1 T286A in *db/db* mice reduced gluconeogenic genes and glycaemia to *db*/+ mouse levels, whereas cyclin D1 K112E caused only minor suppressions of those parameters (Fig. 4b, c and Extended Data Fig. 6e). A similar pattern was observed in mice fed a high-fat diet (Extended Data Fig. 6f–h). We confirmed that cyclin D1 T286A overexpression in *db/db* mice indeed repressed hepatic glucose production by using hyperinsulinemic–euglycemic clamp experiments (Fig. 4d–h).

Our studies support a regulatory model in which insulin facilitates the formation of an active cyclin D1–Cdk4 complex that subsequently suppresses gluconeogenesis, in part by decreasing PGC-1α activity through

©2014 Macmillan Publishers Limited. All rights reserved

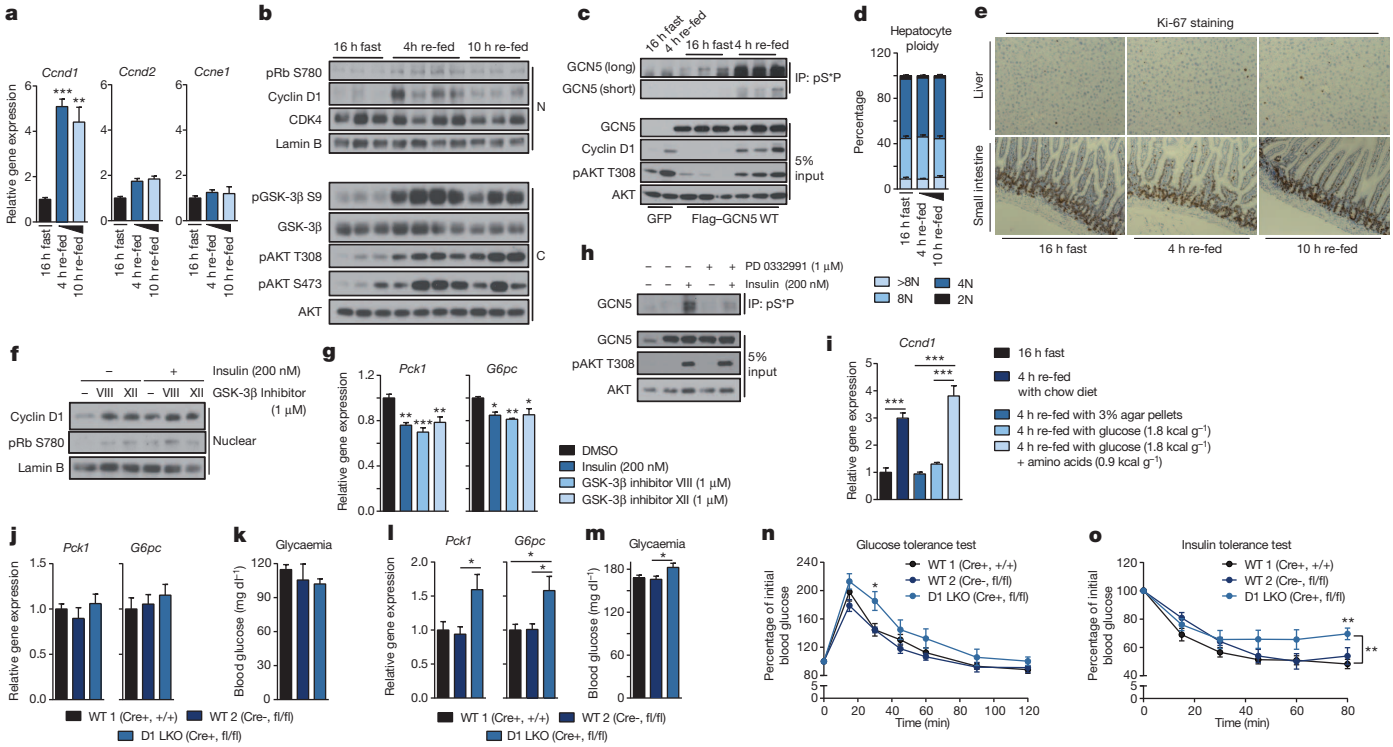


Figure 3 | Cyclin D1–Cdk4 is regulated by insulin/GSK-3β, and hepatic cyclin D1 deletion causes increased gluconeogenesis and glycaemia upon re-feeding. **a**, Cyclin D1 gene expression is increased during re-feeding (one-way ANOVA with Tukey post-test, $n = 3$/fast and10 h re-fed, $n = 4$/4 h re-fed). **b**, Cyclin D1 protein and Rb phosphorylation are induced upon re-feeding (N, nuclear, and C, cytoplasmic, liver extracts). **c**, Phosphorylation of GCN5 is increased upon re-feeding. GCN5 was immunoprecipitated using anti-phosphoS*P (pS*P) antibody from livers infected with GFP or GCN5 adenovirus. **d**, Hepatocyte ploidy does not change upon fasting and re-feeding ($n = 5$). **e**, Ki-67 staining in liver shows no differences upon fasting and re-feeding. Small intestine was used as a positive control. **f**, Nuclear cyclin D1 protein level is increased upon insulin and GSK-3β inhibitor treatment in primary hepatocytes. All cells were infected with cyclin D1 wild-type adenovirus. **g**, Insulin and GSK-3β inhibitors suppress gluconeogenic gene expression. All cells were infected with PGC-1α adenovirus (one-way ANOVA with Tukey post-test, $n = 6$). **h**, Phosphorylation of GCN5 is induced by insulin and blunted by PD 0332991 treatment. GCN5 was overexpressed whereas GFP was used as a negative control. **i**, Cyclin D1 is transcriptionally induced by dietary intake of amino acids. Mice were fasted overnight or re-fed 4 h with chow diet, empty calorie, glucose or glucose and amino-acid diet (one-way ANOVA with Tukey post-test, $n = 5$). **j**, **k**, Liver-specific cyclin D1 LKO mice show no differences on gluconeogenic gene expression and glycaemia during fasting. **l**, **m**, D1 LKO mice show increased gluconeogenic gene expression and glycaemia upon 4 h re-feeding (one-way ANOVA with Tukey post-test, combined four cohorts of $n = 3$/fasting, $n = 5$/re-feeding). **n**, D1 LKO mice show mild glucose intolerance (two-way ANOVA, multiple comparison, $n = 11$/WT1, $n = 7$/WT2, $n = 10$/D1 LKO). **o**, D1 LKO mice show mild insulin intolerance (two-way ANOVA, significant interaction, multiple comparison, $n = 10$/WT1, $n = 7$/WT2, $n = 9$/D1 LKO). Statistical significance is represented by asterisks corresponding to $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Data are shown as mean ± s.e.m.

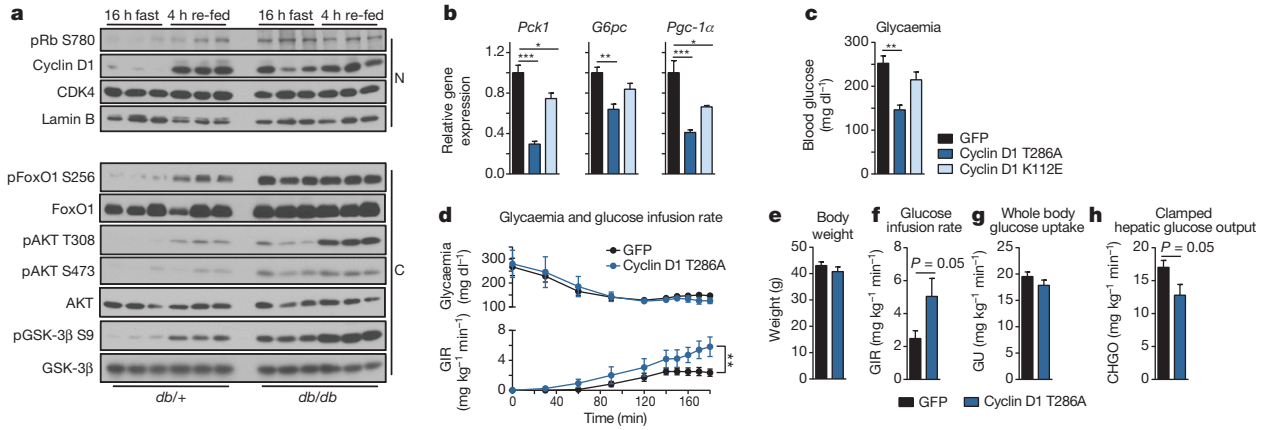


Figure 4 | In diabetic hyperinsulinemic mice, cyclin D1–Cdk4 is dysregulated and hyperactivation of cyclin D1–Cdk4 attenuates the diabetic phenotype. **a**, Cyclin D1 is chronically elevated in livers of *db/db* mice (N, nuclear, and C, cytoplasmic, liver extracts). **b**, **c**, Cyclin D1 T286A, but not cyclin D1 K112E, overexpression in liver represses gluconeogenic genes and glycaemia in *db/db* mice (one-way ANOVA, Tukey post-test, $n = 6$/GFP, D1 K112E, $n = 5$/D1 T286A). **d**–**h**, In *db/db* mice, cyclin D1 T286A overexpression in liver suppresses hepatic glucose production tested by hyperinsulinemic–euglycemic clamp. **d**, Plasma glycaemia and glucose infusion rate (GIR) (two-way ANOVA, significant interaction, $n = 7$). **e**, Body weights. **f**, Clamped glucose infusion rate. **g**, Whole-body glucose uptake (GU). **h**, Clamped hepatic glucose output (CHGO) (average of last 40 min values for **f**, **g**, two-tailed unpaired *t*-test for **d**–**h**, $n = 7$). Statistical significance is represented by asterisks corresponding to $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Data are shown as mean ± s.e.m.

©2014 Macmillan Publishers Limited. All rights reserved

GCN5-mediated acetylation (Extended Data Fig. 6i). Under type 2 diabetic conditions, cyclin D1 protein levels are dysregulated owing to hyperinsulinaemia and insulin resistance, which contributes to its inability to repress gluconeogenesis fully during the fed state. In addition, cyclin D1 is also controlled at the transcriptional level by dietary intake of amino acids, further contributing to insulin's action, although the mechanisms are currently unknown. Despite the functional activation of cyclin D1–Cdk4 during re-feeding, hepatocytes undergo no cell cycle progression. Interestingly, genome-wide association studies have showed that mutations of CDKN2A, a Cdk4 inhibitor, are significantly correlated with a high risk of type 2 diabetes[27]. Combined with the role of cyclin D1–Cdk4 in β-islet pancreatic cell proliferation and the involvement of its target, E2F1, in oxidative metabolism[28,29], induction of gluconeogenesis in liver could predict the tendency of hyperglycaemia in cancer patients treated with Cdk4 inhibitors[30]. Future studies will address the metabolic functions of cell cycle components in different tissues and their use as therapeutic targets in metabolic diseases.

METHODS SUMMARY

The chemical screen was performed with a U-2OS cell line as described in Extended Data Fig. 1a. Immunoprecipitated GCN5 from cells infected with GCN5 adenoviruses was used for *in vitro* acetyltransferase activity assay. Glutathione *S*-transferase (GST)–cyclin D1–Cdk4, GST–Rb and 6 × His-GCN5 recombinant proteins were purchased whereas others were generated in the laboratory. Glucose levels in culture medium were measured with primary hepatocytes overexpressing PGC-1α. All animal studies were performed with 8- to 10-week-old male C57BL/6 mice unless indicated otherwise. PD 0332991 (150 mg kg^{-1}, pH 4) was administered by oral gavage. Adenoviruses were delivered by tail-vein injection with dosages of 1.5×10^9 to 2.0×10^9 infectious particles per animal. Pyruvate tolerance test was performed by intraperitoneal injection of pyruvate (2 g kg^{-1}) to overnight fasted mice whereas glucose tolerance (1 g kg^{-1}) and insulin tolerance tests (0.75 U kg^{-1}, diluted in PBS) were performed with mice fasted for 6 h. Diets containing empty calorie, glucose or glucose and amino acids were formulated in the laboratory. High-fat diet (60 kcal%) was administered to animals for 3 months before experiments. Detailed experimental procedures can be found in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 February 2013; accepted 20 March 2014.
Published online 25 May 2014.

1. Nakae, J., Park, B. C. & Accili, D. Insulin stimulates phosphorylation of the forkhead transcription factor FKHR on serine 253 through a wortmannin-sensitive pathway. *J. Biol. Chem.* **274,** 15982–15985 (1999).
2. Cho, H. *et al.* Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKBβ). *Science* **292,** 1728–1731 (2001).
3. Dentin, R. *et al.* Insulin modulates gluconeogenesis by inhibition of the coactivator TORC2. *Nature* **449,** 366–369 (2007).
4. Matsumoto, M., Pocai, A., Rossetti, L., Depinho, R. A. & Accili, D. Impaired regulation of hepatic glucose production in mice lacking the forkhead transcription factor Foxo1 in liver. *Cell Metab.* **6,** 208–216 (2007).
5. Puigserver, P. *et al.* Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1α interaction. *Nature* **423,** 550–555 (2003).
6. Li, X., Monks, B., Ge, Q. & Birnbaum, M. J. Akt/PKB regulates hepatic metabolism by directly inhibiting PGC-1α transcription coactivator. *Nature* **447,** 1012–1016 (2007).
7. Estall, J. L. *et al.* Sensitivity of lipid metabolism and insulin signaling to genetic alterations in hepatic peroxisome proliferator-activated receptor-γ coactivator-1α expression. *Diabetes* **58,** 1499–1508 (2009).
8. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1α and SIRT1. *Nature* **434,** 113–118 (2005).
9. Lerin, C. *et al.* GCN5 acetyltransferase complex controls glucose metabolism through transcriptional repression of PGC-1α. *Cell Metab.* **3,** 429–438 (2006).
10. McKeehan, W. L., Adams, P. S. & Rosser, M. P. Direct mitogenic effects of insulin, epidermal growth factor, glucocorticoid, cholera toxin, unknown pituitary factors and possibly prolactin, but not androgen, on normal rat prostate epithelial cells in serum-free, primary cell culture. *Cancer Res.* **44,** 1998–2010 (1984).
11. Ish-Shalom, D. *et al.* Mitogenic properties of insulin and insulin analogues mediated by the insulin receptor. *Diabetologia* **40** (suppl 2), S25–S31 (1997).
12. Soni, R. *et al.* Inhibition of cyclin-dependent kinase 4 (Cdk4) by fascaplysin, a marine natural product. *Biochem. Biophys. Res. Commun.* **275,** 877–884 (2000).
13. Matsushime, H. *et al.* Identification and properties of an atypical catalytic subunit (p34PSK-J3/cdk4) for mammalian D type G1 cyclins. *Cell* **71,** 323–334 (1992).
14. Kato, J., Matsushime, H., Hiebert, S. W., Ewen, M. E. & Sherr, C. J. Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes Dev.* **7,** 331–342 (1993).
15. Fry, D. W. *et al.* Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts. *Mol. Cancer Ther.* **3,** 1427–1438 (2004).
16. Boyer, L. A. *et al.* Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol. Cell* **10,** 935–942 (2002).
17. Rodgers, J. T. & Puigserver, P. Fasting-dependent glucose and lipid metabolic response through hepatic sirtuin 1. *Proc. Natl Acad. Sci. USA* **104,** 12861–12866 (2007).
18. Diehl, J. A., Zindy, F. & Sherr, C. J. Inhibition of cyclin D1 phosphorylation on threonine-286 prevents its rapid degradation via the ubiquitin-proteasome pathway. *Genes Dev.* **11,** 957–972 (1997).
19. Hinds, P. W., Dowdy, S. F., Eaton, E. N., Arnold, A. & Weinberg, R. A. Function of a human cyclin gene as an oncogene. *Proc. Natl Acad. Sci. USA* **91,** 709–713 (1994).
20. Anders, L. *et al.* A systematic screen for CDK4/6 substrates links FOXM1 phosphorylation to senescence suppression in cancer cells. *Cancer Cell* **20,** 620–634 (2011).
21. Abella, A. *et al.* Cdk4 promotes adipogenesis through PPARgamma activation. *Cell Metab.* **2,** 239–249 (2005).
22. Diehl, J. A., Cheng, M., Roussel, M. F. & Sherr, C. J. Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev.* **12,** 3499–3511 (1998).
23. Choi, Y. J. *et al.* The requirement for cyclin D function in tumor maintenance. *Cancer Cell* **22,** 438–451 (2012).
24. Pinkert, C. A., Ornitz, D. M., Brinster, R. L. & Palmiter, R. D. An albumin enhancer located 10 kb upstream functions along with its promoter to direct efficient, liver-specific expression in transgenic mice. *Genes Dev.* **1,** 268–276 (1987).
25. Macotela, Y. *et al.* Dietary leucine–an environmental modifier of insulin resistance acting on multiple levels of metabolism. *PLoS ONE* **6,** e21187 (2011).
26. Wang, Y. *et al.* Inositol-1,4,5-trisphosphate receptor regulates hepatic gluconeogenesis in fasting and diabetes. *Nature* **485,** 128–132 (2012).
27. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316,** 1336–1341 (2007).
28. Rane, S. G. *et al.* Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in β-islet cell hyperplasia. *Nature Genet.* **22,** 44–52 (1999).
29. Blanchet, E. *et al.* E2F transcription factor-1 regulates oxidative metabolism. *Nature Cell Biol.* **13,** 1146–1152 (2011).
30. Schwartz, G. K. *et al.* Phase I study of PD 0332991, a cyclin-dependent kinase inhibitor, administered in 3-week cycles (Schedule 2/1). *Br. J. Cancer* **104,** 1862–1868 (2011).

Acknowledgements We thank all the members of the Puigserver laboratory for discussions and suggestions about this project. We also appreciate the consultations with and efforts by M. Jedrychowski and S. Gygi for proteomic analysis. Y.L. was supported in part by a 21st Century Leaders scholarship from Ewha Womans University. J.E.D. was supported in part by a National Research Service Award Kirschstein Fellowship from the National Institutes of Health (NIH). The participating researchers were supported with funds from the Dana-Farber Cancer Institute and with grants from the American Diabetes Association, Department of Defense, NIH/National Institute of Diabetes and Digestive and Kidney Diseases (RO1069966 and R24DK080261-06), NIH (RO3 MH092174) awarded to P.P., NIH (RO1 CA108420) awarded to P.S. and NIH (DK059635) awarded to Yale's Mouse Metabolic Phenotyping Center/G.I.S.

Author Contributions P.P. and Y.L. conceived the project. Y.L. designed and performed most of the experiments aided by discussions with P.P. J.E.D. also contributed to the design and discussions of the manuscript. Additional advice in the design and execution of the experiments was provided by F.V. and J.-H.L. H.C. assisted with mice experiments. Y.J.C. and P.S. provided D1 fl/fl mice and helped with BrdU incorporation and Ki-67 staining experiments. M.J., J.P.C., H.-B.R., X.Y. and G.I.S. performed hyperinsulinemic–euglycemic clamp experiments, and analysed, interpreted and discussed the data. N.T. assisted with performance and analysis of chemical screening. Y.L. and P.P. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.P. (pere_puigserver@dfci.harvard.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13269

Structure of a lipid–bound extended synaptotagmin indicates a role in lipid transfer

Curtis M. Schauder[1], Xudong Wu[1]*, Yasunori Saheki[1,2]*, Pradeep Narayanaswamy[3], Federico Torta[4], Markus R. Wenk[4], Pietro De Camilli[1,2] & Karin M. Reinisch[1]

Growing evidence suggests that close appositions between the endoplasmic reticulum (ER) and other membranes, including appositions with the plasma membrane (PM), mediate exchange of lipids between these bilayers. The mechanisms of such exchange, which allows lipid transfer independently of vesicular transport, remain poorly understood. The presence of a synaptotagmin-like mitochondrial-lipid-binding protein (SMP) domain, a proposed lipid-binding module, in several proteins localized at membrane contact sites has raised the possibility that such domains may be implicated in lipid transport[1,2]. SMP-containing proteins include components of the ERMES complex, an ER–mitochondrial tether[3], and the extended synaptotagmins (known as tricalbins in yeast), which are ER–PM tethers[4–6]. Here we present at 2.44 Å resolution the crystal structure of a fragment of human extended synaptotagmin 2 (E-SYT2), including an SMP domain and two adjacent C2 domains. The SMP domain has a β-barrel structure like protein modules in the tubular-lipid-binding (TULIP) superfamily. It dimerizes to form an approximately 90-Å-long cylinder traversed by a channel lined entirely with hydrophobic residues, with the two C2A–C2B fragments forming arched structures flexibly linked to the SMP domain. Importantly, structural analysis complemented by mass spectrometry revealed the presence of glycerophospholipids in the E-SYT2 SMP channel, indicating a direct role for E-SYTs in lipid transport. These findings provide strong evidence for a role of SMP-domain-containing proteins in the control of lipid transfer at membrane contact sites and have broad implications beyond the field of ER-to-PM appositions.

ER–PM contact sites, which are present in all eukaryotes, are thought to have multiple functions[7–9], including an evolutionarily conserved role in the regulation of lipid composition and metabolism at the PM[8–11]. E-SYTs are ER-resident proteins that act as tethers linking the ER and PM membranes into close apposition[4–6]. There are three E-SYTs (E-SYT1, 2 and 3) in mammalian cells, which homo- or heterodimerize[6]. Each E-SYT has an amino-terminal ER-membrane anchor, followed by a short linker region, an SMP domain, and three or more C2 domains[6] (Fig. 1a). The ER-membrane anchor and C2 domains are required for tethering, whereas the role of the SMP domain has been unclear[6]. Bioinformatics analyses suggesting that SMP domains belong to the TULIP superfamily were therefore intriguing because several TULIP proteins are known to bind lipids and participate in lipid transfer[1,2].

To obtain insights regarding E-SYT function, we determined the crystal structure of a fragment of human E-SYT2 (residues 163–634) comprising the SMP, C2A and C2B domains (Extended Data Table 1). The carboxy-terminal portion of E-SYT2, which consists of an extended linker region followed by a third C2 domain (C2C) that interacts with phosphatidylinositol 4,5-bisphosphate ($PI(4,5)P_2$) at the PM[6], was not in the construct.

This E-SYT2 fragment dimerizes both in solution, as assessed by size-exclusion chromatography, and in the crystal. Each SMP domain is a β-barrel (Fig. 1b), comprising a highly twisted β-sheet and two α-helices, H1and H3. A third helix, H2, partially caps one end of the barrel. The other end of the barrel associates with the corresponding end of the second SMP domain to form a 90-Å-long cylinder. The residues at the interface of the two SMP domains are among the most highly conserved (Extended Data Fig. 1), consistent with the physiological relevance of the dimer, and they probably mediate homo- and heterodimerization of E-SYTs[6]. An approximately 10-Å-diameter channel lined exclusively with hydrophobic residues runs through the cylinder (Fig. 1c). It connects with solvent at both ends and via a narrow seam along the length of the cylinder.

There is no significant similarity between the primary sequences of the E-SYT2 SMP domain and proteins of known fold. Nevertheless, as

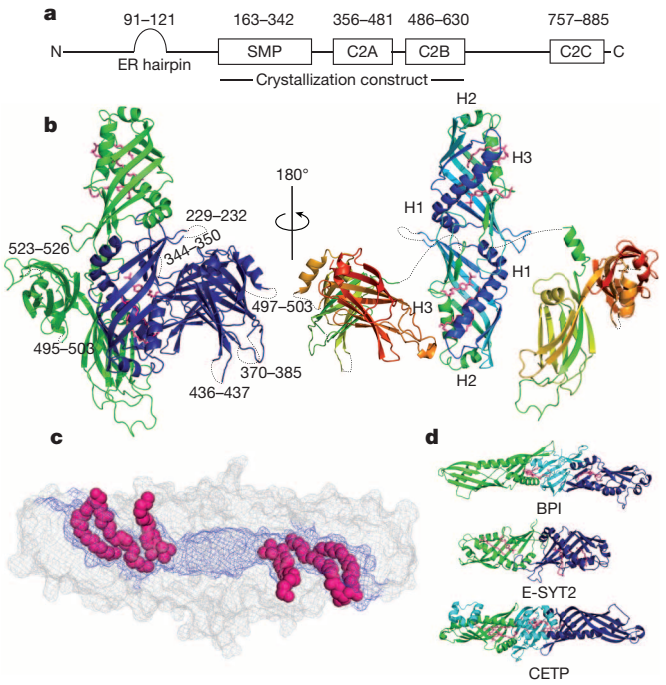


Figure 1 | E-SYT2 structure. **a**, E-SYT2 domain architecture. Domain boundaries are indicated as residue numbers. **b**, Left, monomers in the E-SYT2 dimer are green or blue. C2A–C2B arches (see also Fig. 3) contact the SMP barrel as in the crystal. Right, E-SYT2 is coloured from blue (N terminus) to red (C terminus), and the C2A–C2B arches are pulled away from the SMP dimer. The 'seam' along the side of the SMP dimer is partially visible. Detergent and lipid molecules are pink. **c**, Surface representation of the SMP dimer. Hydrophobic residues (blue) line the channel. Lipid fatty acid moieties are in space-filling representation. **d**, Ribbon diagrams for BPI (Protein Data Bank (PDB) accession 1BP1), E-SYT2 and CTEP (PDB accession 2OBD). Structural elements in BPI or CTEP not present in E-SYT2 are cyan.

[1]Department of Cell Biology, Yale School of Medicine, New Haven, Connecticut 06520, USA. [2]Program in Cellular Neuroscience, Neurodegeneration, and Repair, Yale School of Medicine, New Haven, Connecticut 06510, USA; and Howard Hughes Medical Institute, Yale School of Medicine, New Haven, Connecticut 06510, USA. [3]Department of Biological Sciences, National University of Singapore, 117543 Singapore. [4]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 117599 Singapore.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

predicted by bioinformatics studies[1,2], our structure firmly establishes the SMP domain as a member of the TULIP superfamily. In previously characterized TULIP proteins such as bactericidal/permeability-increasing protein (BPI) and cholesteryl ester transfer protein (CETP) (Fig. 1d), SMP-like modules also form tubular structures harbouring an elongated hydrophobic channel[12,13]. BPI and CETP do not dimerize, however, as each monomer comprises two SMP-like domains rather than one, with the two modules arranged in tandem and connected by a β-sheet not present in E-SYTs.

Within the SMP channel, the electron density maps show clear densities for two lipid-like molecules per E-SYT2 monomer (Fig. 2a, b). One density is consistent with a diacylglycerol lipid, whereas the other is reminiscent of Triton X-100, a detergent used in protein purification. Triton X-100 was modelled into this density with its 4-(1,1,3,3 tetramethylbutyl)-phenyl group buried within the hydrophobic channel and its hydrophilic polyethylene chain extended through the seam into the solvent (Fig. 2). A non-denaturing mass spectrometry (MS) analysis[14] of recombinant E-SYT2 purified in the absence of Triton X-100 shows that ligand-bound proteins are present and that two lipid molecules can bind per monomer (Extended Data Fig. 2), so probably a second lipid molecule binds in place of the detergent in a physiological context. Denaturing negative-ionization high-resolution MS and MS/MS analysis further identified ligands of the recombinant E-SYT2 as two major components of the bacterial membrane, phosphatidylglycerol and phosphatidylethanolamine (PE) (Extended Data Fig. 3 and Extended Data Table 2).

The electron density maps show that the lipid fatty acid moieties lie in the hydrophobic channel, whereas the polar head group protrudes through the seam and is solvated. There is no density for the head group, implying that it does not make extensive contacts with E-SYT2 and is disordered. This suggests that the SMP domain could bind a variety of glycerophospholipids. BPI and CETP bind diverse lipids in a similar manner, with hydrophobic portions buried in the hydrophobic channel and the hydrophilic groups exposed to solvent along the seam[12,13].

To identify lipids bound to E-SYT2 in mammalian cells, a 3×Flag-tagged E-SYT2 fragment comprising the SMP and C2A–C2B domains was expressed in Expi293 cells and purified by anti-Flag immunoprecipitation. Denaturing MS of this construct and quantitative lipidomic analysis revealed the presence of the glycerophospholipids phosphatidylcholine (PC; 67%), PE (21%), phosphatidylinositol (7%) and phosphatidylserine (3%) (Extended Data Figs 4, 5 and Extended Data Table 3). These relative amounts of the major phospholipids roughly reflect their relative abundance in animal cell membranes[15], suggesting a lack of specificity with respect to particular glycerophospholipid ligands. We looked for but did not detect other lipid species such as lysophosphatidylcholine, sphingosine or sphingosine-1-phosphate, sphingomyelin,

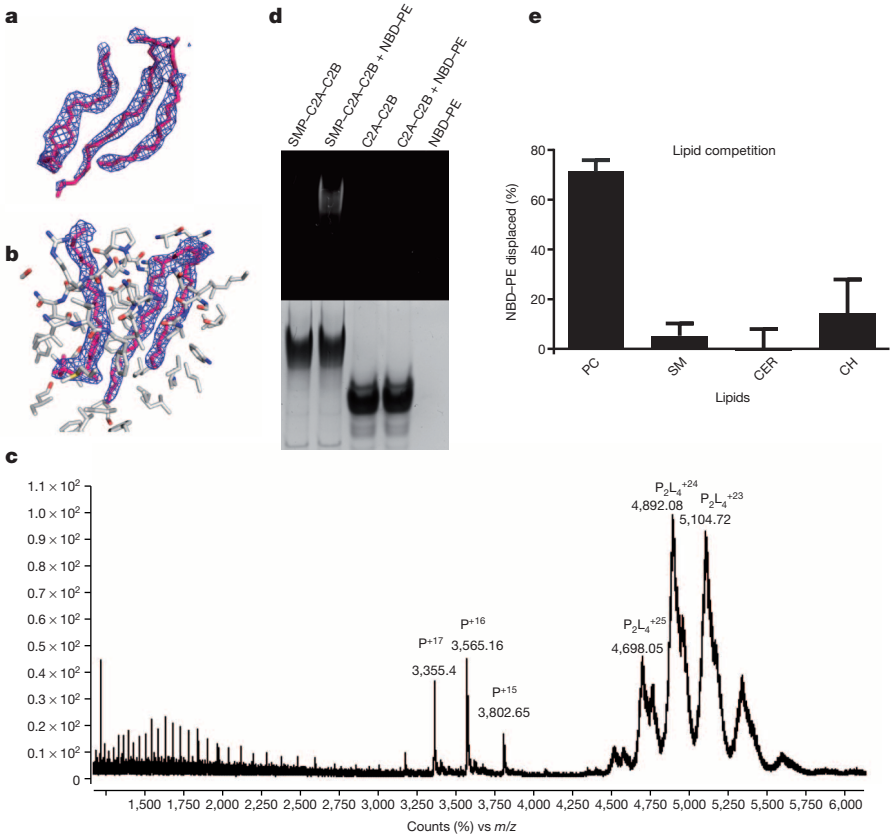


Figure 2 | E-SYT2 binds lipid-like molecules in the SMP hydrophobic channel. **a**, $2F_o - F_c$ electron density (1.0σ) into which the Triton X-100 detergent and phosphoglycerol lipid were modelled; models after refinement are superimposed. Lipid head groups were not modelled. **b**, Final $2F_o - F_c$ electron density for lipid-like molecules. Protein atoms within 8 Å of lipid are indicated. **c**, Positive mode, non-denaturing electrospray ionization (ESI) mass spectrum of E-SYT2 expressed in mammalian cells. In native conditions (20 mM ammonium acetate, pH 6.9) a conformation of apo E-SYT2 centred around the +16 charge state is the most abundant monomeric state (indicated as P); molecular weight = 57,032.4 ± 1.6 Da. The most represented species in native conditions are dimers containing at least four lipid molecules (P_2L_4) with a charge state centred around +24. *m*/*z*, mass to charge ratio. **d**, An E-SYT2 construct that includes the SMP domain binds fluorescent NBD–PE, whereas the C2A–C2B arch alone does not. Proteins were incubated with fluorescent lipid, then analysed by native PAGE, which was examined for fluorescence (top) and stained by Coomassie (bottom). The band with the SMP-containing protein co-migrates with fluorescent PE. **e**, E-SYT2 preloaded with NBD–PE, then further purified, was incubated with phosphatidylcholine (PC), sphingomyelin (SM), ceramide (CER) or cholesterol (CH). Samples were visualized on native gels by fluorescence and Coomassie staining. Fluorescence signal was normalized to the amount of protein in each lane, which was assessed by Coomassie staining. The experiment was performed five times; error bars represent standard deviation. PC, but not any of the sterols, displaces NBD–PE. (See also Extended Data Fig. 7.)

©2014 Macmillan Publishers Limited. All rights reserved

any phosphoinositides, or notable amounts of ceramide. Nor were cholesterol esters, which are bound by CETP[13], or cholesterol detected.

When analysed by non-denaturing MS, E-SYT2 was found to occur both as monomer and dimer (Fig. 2c). The charge state distribution centred around +16 represents monomers mainly in the apo form, corresponding to a molecular weight of 57,030 dalton (Da) after deconvolution (Extended Data Fig. 6a). Only a low percentage of the monomers (~10%) were bound to one or two lipid molecules, with masses corresponding to the main lipid species identified previously. Most of the protein was dimeric with a charge state distribution centred around +24. After deconvolution of this spectrum (Extended Data Figs 6b, c and 7), the molecular weight of the dimer is ~117,500 Da, suggesting the presence of at least four bound lipid molecules (assuming an average mass of 750–800 Da per lipid). Almost no dimers were detected in the apo form. These findings indicate that dimerization is required for stable binding of the lipids.

Binding to glycerophospholipids was also demonstrated *in vitro* by testing the ability of purified lipids to displace preloaded fluorescent 7-nitrobenz-2-oxa-1,3-diazol-4-yl–1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (NBD–PE) from E-SYT2 (Fig. 2d, e). Incubation of preloaded E-SYT2 with PC displaced NBD–PE, as revealed by loss of fluorescence in the band corresponding to E-SYT2 in native polyacrylamide gel electrophoresis (PAGE; Fig. 2e). In contrast, consistent with the MS results, cholesterol, sphingomyelin and ceramide did not displace PE, suggesting that they are not bound by E-SYT2 even when abundantly present. The greater rigidity of these lipids as compared to glycerophospholipids may prevent their entry into the E-SYT2 lipid-binding site.

C2A and C2B both fold into type II C2 domains and are rigidly linked into an arch spanning ~70 Å across (Fig. 3 shows best view of arch), as recently described[16]. In the crystal structure, both C2A–C2B modules of the E-SYT2 dimer contact the same SMP domain, with different C2A–C2B residues interacting with different patches on the SMP (Fig. 1a, left). Residues at the interfaces are not conserved (Extended Data Fig. 1), suggesting that the C2–SMP interactions are crystal contacts and that C2A–C2B does not interact with the SMP dimer in solution (Fig. 1b, right). Rather, the E-SYT2 C2A–C2B fragment interacts with lipid bilayers in a Ca^{2+}-dependent manner[17]. Accordingly, calcium-coordinating residues were identified in loops at the end of the arch structure in C2A[16].

The distance between the ER and PM membranes at contact sites is estimated at 100–300 Å (refs 8, 18, 19). E-SYT2 spans this space with an N-terminal membrane anchor inserted into the ER and its C2C domain bound to the PM via an interaction with $PI(4,5)P_2$, which is enriched there[6] (Fig. 3). Although C2C, when expressed alone, is recruited to the PM[6,17], we and others[17] did not observe any association of C2A–C2B with a particular organelle in living cells. Given the geometric restraints imposed by domain dimensions and linker lengths, an interaction between C2A–C2B and the cytosolic surface of the ER is plausible and consistent with its previously observed acidic-phospholipid-independent membrane interactions[17]. An interaction between C2A–C2B and the PM at very narrow ER–PM contacts (less than ~100 Å), where it can reach the PM, is also plausible, however.

The structure of the E-SYT2 SMP domain and identification of lipids in its hydrophobic channel conclusively establishes this domain as a lipid-binding module. The structural similarity of E-SYT2 to proteins in the TULIP superfamily, several of which function in lipid transfer extra-cytosolically[2,12,13], further suggests a role in lipid transport. Moreover, the presence of this module in proteins that localize to contacts between the ER and other membranes[4–6,20] strongly indicates a role in the transfer of lipids from their site of synthesis in the ER to these membranes, such as the PM in the case of E-SYT2. In answer to how such a transfer may take place, SMP dimers could in principle form a bridge between apposed membranes along which lipids are transferred ('tunnel model'; Fig. 3), with their hydrophobic moieties sliding along the hydrophobic channel and their hydrophilic head groups solvent exposed at the seam. CETP was proposed to function in exactly this manner to channel cholesterol esters between lipoprotein particles[21]. However, as it is only 90 Å long, the E-SYT2 SMP dimer may be too short to span the inter-membrane space. Another possibility, plausible for ER–PM distances under ~200 Å, is that the SMP dimer shuttles between membranes, extracting lipid from one and delivering it to the second ('shuttle model'; Fig. 3). In a third, not mutually exclusive model, E-SYT2 functions in concert with other lipid transfer proteins to effect transport. This model is appealing because it does not require E-SYTs preferentially

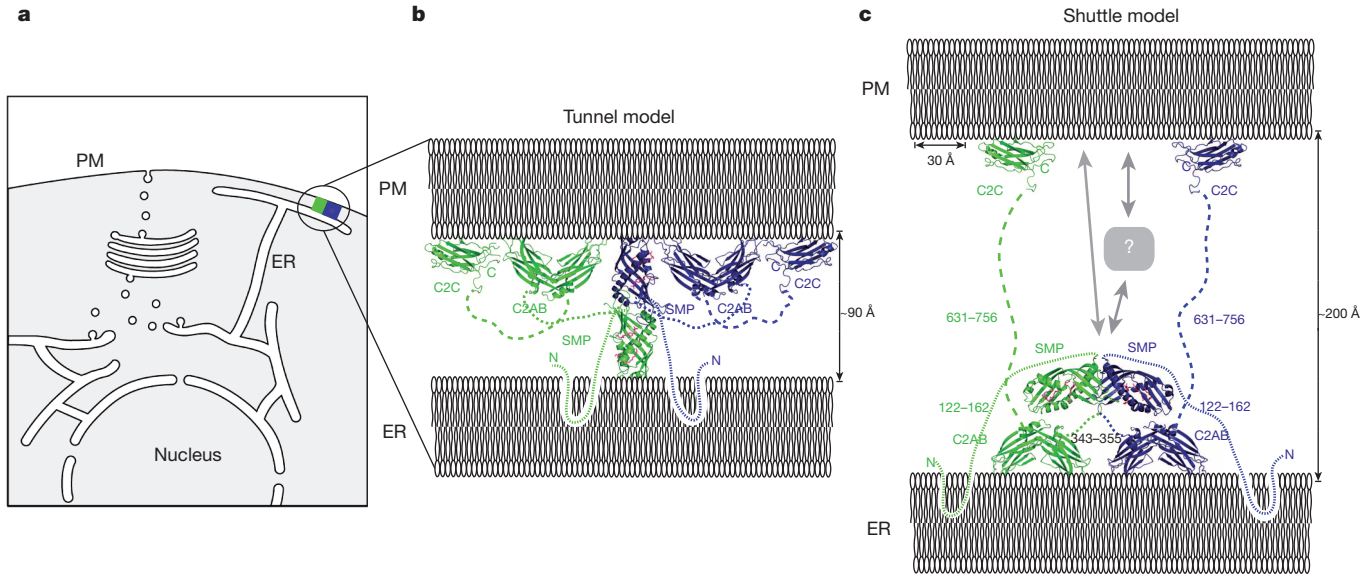


Figure 3 | ER–PM contact site and a model of E-SYT2. **a**, Schematic of a cell with an ER–PM contact site. **b**, 'Tunnel model' of E-SYT2, in which E-SYT2 bridges the ER and PM to transfer lipids between them. The E-SYT2 SMP dimer is ~90 Å long, so the ER and PM cannot be farther apart if E-SYT2 acts as a tunnel. E-SYT2 is anchored to the ER via its N-terminal region and to the PM via the C2C domain (PDB accession 2DMG). C2A–C2B interacts with membrane via loops at the ends of the arch and could interact with the ER, PM or either, depending on the inter-membrane distance. The model is to scale. **c**, 'Shuttle model' of E-SYT2 tethering the ER and PM and drawn to scale. Length of unstructured linker regions is indicated in residues. At larger ER–PM distances, as shown here, C2A–C2B (C2AB) cannot reach the PM. The SMP dimer may shuttle lipids between the ER and PM, potentially together with partner proteins (labelled with a question mark).

©2014 Macmillan Publishers Limited. All rights reserved

to bind specific lipids, which could instead be selected by the partner proteins. Phosphatidylinositol transfer proteins (PITPs) and proteins in the OSH family are interesting candidates because they can interact with specific lipids[22,23] and are thought to function in lipid transfer at contact sites[8,9].

The identification of E-SYTs as lipid transfer proteins represents an important step forward in dissecting the biology of ER–PM contact sites. Nevertheless, our work is also more generally applicable to other contact sites, such as ER–mitochondrial contacts, where the ERMES tethering complex is predicted to contain three SMP modules[1,2]. We suspect that these hetero-dimerize in a head-to-head fashion, as observed in the E-SYT2 SMP homodimer, to form one or more lipid carrier modules with hydrophobic lipid-binding channels.

METHODS SUMMARY

Native or selenomethionine-substituted recombinant human E-SYT2 (residues 163–634) were overexpressed in and purified from *Escherichia coli*. The protein was crystallized by the hanging-drop method, equilibrating against mother liquor comprising 100 mM malonic acid, imidazole and boric acid (MIB) buffer at pH 4.3–4.6 and 12.5–14% PEG 1500, and additionally 0.1 M NaI for the selenomethionine-substituted protein. We used the multi-wavelength anomalous dispersion method with the selenomethionine-substituted crystals for phasing (Extended Data Table 1). The structure was refined against data from native crystals (Extended Data Table 1). Structure determination, MS analysis and *in vitro* lipid-binding assays are described in more detail in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 October 2013; accepted 21 March 2014.

Published online 18 May 2014.

1. Kopec, K. O., Alva, V. & Lupas, A. N. Homology of SMP domains to the TULIP superfamily of lipid-binding proteins provides a structural basis for lipid exchange between ER and mitochondria. *Bioinformatics* **26,** 1927–1931 (2010).
2. Kopec, K. O., Alva, V. & Lupas, A. N. Bioinformatics of the TULIP domain superfamily. *Biochem. Soc. Trans.* **39,** 1033–1038 (2011).
3. Kornmann, B. *et al.* An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science* **325,** 477–481 (2009).
4. Toulmay, A. & Prinz, W. A. A conserved membrane binding domain targets proteins to organelle contact sites. *J. Cell Sci.* **125,** 49–58 (2012).
5. Manford, A. G., Stefan, C. J., Yuan, H. L., MacGurn, J. A. & Emr, S. D. ER-to-plasma membrane tethering proteins regulate cell signaling and ER morphology. *Dev. Cell* **23,** 1129–1140 (2012).
6. Giordano, F. *et al.* PI(4,5)P_2-Dependent and Ca^{2+}-regulated ER-PM interactions mediated by the extended synaptotagmins. *Cell* **153,** 1494–1509 (2013).
7. Levine, T. & Loewen, C. J. Inter-organelle membrane contact sites: through a glass, darkly. *Curr. Opin. Cell Biol.* **18,** 371–378 (2006).
8. Toulmay, A. & Prinz, W. A. Lipid transfer and signaling at organelle contact sites: the tip of the iceberg. *Curr. Opin. Cell Biol.* **23,** 458–463 (2011).
9. Stefan, C. J., Manford, A. G. & Emr, S. D. ER-PM connections: sites of information transfer and inter-organelle communication. *Curr. Opin. Cell Biol.* **25,** 434–442 (2013).
10. Stefan, C. J. *et al.* Osh proteins regulate phosphoinositide metabolism at ER-plasma membrane contact sites. *Cell* **144,** 389–401 (2011).
11. Tavassoli, S. *et al.* Plasma membrane-endoplasmic reticulum contact sites regulate phosphatidylcholine synthesis. *EMBO Rep.* **14,** 434–440 (2013).
12. Beamer, L. J., Carroll, S. F. & Eisenberg, D. Crystal structure of human BPI and two bound phospholipids at 2.4 Angstrom resolution. *Science* **276,** 1861–1864 (1997).
13. Qiu, X. *et al.* Crystal structure of cholesteryl ester transfer protein reveals a long tunnel and four bound lipid molecules. *Nature Struct. Mol. Biol.* **14,** 106–113 (2007).
14. Krylova, I. N. *et al.* Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* **120,** 343–355 (2005).
15. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature Rev. Mol. Cell Biol.* **9,** 112–124 (2008).
16. Xu, J. *et al.* Structure and Ca^{2+}-binding properties of the tandem C_2 domains of E-Syt2. *Structure* **22,** 269–280 (2013).
17. Min, S.-W., Chang, W.-P. & Südhof, T. C. E-Syts, a family of membranous Ca^{2+}-sensor proteins with multiple C_2 domains. *Proc. Natl Acad. Sci. USA* **104,** 3823–3828 (2007).
18. Orci, L. *et al.* STIM1-induced precortical and cortical subdomains of the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA* **106,** 19358–19362 (2009).
19. West, M., Zurek, N., Hoenger, A. & Voeltz, G. K. A 3D analysis of yeast ER structure reveals how ER domains are organized by membrane curvature. *J. Cell Biol.* **193,** 333–346 (2011).
20. Helle, S. C. J. *et al.* Organization and function of membrane contact sites. *Biochim. Biophys. Acta* **1833,** 2526–2541 (2013).
21. Zhang, L. *et al.* Structural basis of transfer between lipoproteins by cholesteryl ester transfer protein. *Nature Chem. Biol.* **8,** 342–349 (2012).
22. de Saint-Jean, M. *et al.* Osh4p exchanges sterols for phosphatidylinositol 4-phosphate between lipid bilayers. *J. Cell Biol.* **195,** 965–978 (2011).
23. Schulz, T. A. *et al.* Lipid-regulated sterol transfer between closely apposed membranes by oxysterol-binding protein homologues. *J. Cell Biol.* **187,** 889–903 (2009).

Acknowledgements We are grateful to the staff at NE-CAT beamline 24ID-C at the Advanced Photon Source for assistance with data collection. We thank H. Shen and N. Kumar for discussion. This work was supported by the National Institutes of Health (GM080616 to K.M.R., R37NS36251 and DK082700 to P.D.C.). M.R.W. and F.T. are funded by grants from the National University of Singapore via the Life Sciences Institute (LSI) and the Biochemical Research Council (BMRC-SERC Diag-034). C.M.S. is recipient of an National Science Foundation Graduate Research Fellowship (DGE-1122492). Y.S. was supported by the Uehara Memorial Foundation and the Japanese Society for Promotion of Science.

Author Contributions All authors participated in the design of the experiments, data analysis, and manuscript editing. C.M.S., X.W., P.N., F.T. and Y.S. carried out the experiments. F.T., M.R.W., P.D.C. and K.M.R. wrote this manuscript.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under accession number 4P42. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M.R. (karin.reinisch@yale.edu) or P.D.C. (pietro.decamilli@yale.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13295

BRCA1 controls homologous recombination at Tus/*Ter*-stalled mammalian replication forks

Nicholas A. Willis[1], Gurushankar Chandramouly[1]†, Bin Huang[1]†, Amy Kwok[1]†, Cindy Follonier[2], Chuxia Deng[3] & Ralph Scully[1]

Replication fork stalling can promote genomic instability, predisposing to cancer and other diseases[1–3]. Stalled replication forks may be processed by sister chromatid recombination (SCR), generating error-free or error-prone homologous recombination (HR) outcomes[4–8]. In mammalian cells, a long-standing hypothesis proposes that the major hereditary breast/ovarian cancer predisposition gene products, BRCA1 and BRCA2, control HR/SCR at stalled replication forks[9]. Although BRCA1 and BRCA2 affect replication fork processing[10–12], direct evidence that *BRCA* gene products regulate homologous recombination at stalled chromosomal replication forks is lacking, due to a dearth of tools for studying this process. Here we report that the *Escherichia coli* Tus/*Ter* complex[13–16] can be engineered to induce site-specific replication fork stalling and chromosomal HR/SCR in mouse cells. Tus/*Ter*-induced homologous recombination entails processing of bidirectionally arrested forks. We find that the Brca1 carboxy (C)-terminal tandem BRCT repeat and regions of Brca1 encoded by exon 11—two Brca1 elements implicated in tumour suppression—control Tus/*Ter*-induced homologous recombination. Inactivation of either Brca1 or Brca2 increases the absolute frequency of 'long-tract' gene conversions at Tus/*Ter*-stalled forks, an outcome not observed in response to a site-specific endonuclease-mediated chromosomal double-strand break. Therefore, homologous recombination at stalled forks is regulated differently from homologous recombination at double-strand breaks arising independently of a replication fork. We propose that aberrant long-tract homologous recombination at stalled replication forks contributes to genomic instability and breast/ovarian cancer predisposition in *BRCA* mutant cells.

Tus binds the 23 base pair (bp) *Ter* site to induce polar replication fork arrest in *E. coli*[13–16]. To determine whether Tus/*Ter* can arrest mammalian replisomes, we introduced six *TerB* sites into a plasmid containing the Epstein–Barr virus nuclear antigen 1 (EBNA1)-binding origin of replication (p6×*Ter*Ori, Fig. 1a). EBNA1 recruits mammalian replication factors, mediating predominantly unidirectional plasmid replication, due to a replication block at EBNA1-bound FR (family of repeats)[17]. In p6×*Ter*Ori, the major clockwise fork approaches the 'non-permissive' (fork-stalling) face of Tus/*Ter* (Fig. 1a). We used two-dimensional DNA gel electrophoresis with Southern blotting[18] to visualize replication through 6×*Ter*. Transfection of 293E cells, which express EBNA1 (ref. 19), with p6×*Ter*Ori and control empty vector revealed plasmid replication intermediates (arc A, Fig. 1b). Co-transfection of p6×*Ter*Ori and myc-tagged Tus revealed site-specific stalling of the clockwise fork (spot B, Fig. 1b, c and Extended Data Fig. 1). TusH144A, a *Ter*-binding-impaired mutant[16], induced minimal fork stalling. Reversal of 6×*Ter* to the 'permissive' orientation (6×REV*Ter*, Fig. 1b) also supported Tus-dependent stalling of the clockwise fork, albeit less efficiently than non-permissive 6×*Ter* (Fig. 1b, c). The FR/EBNA1 replication block is incomplete[20]. A weaker Tus/*Ter*-dependent double-Y spot (C, Fig. 1b and Extended Data Fig. 1) reflects bidirectional fork arrest at 6×*Ter*. We estimate the FR/EBNA1 and Tus/6×*Ter* replication block efficiencies as ~70% (Extended

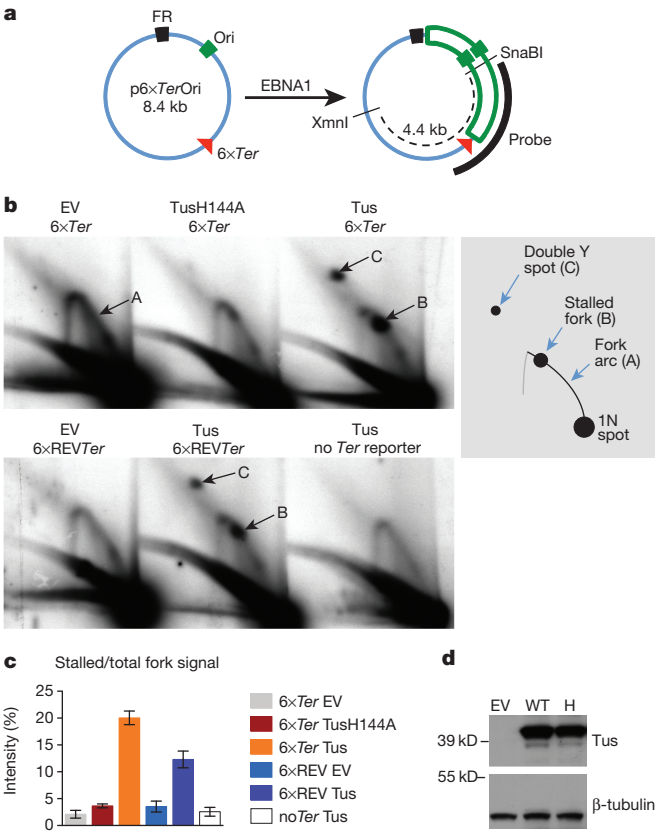


Figure 1 | Tus/*Ter*-induced replication fork stalling in mammalian cells. **a**, EBNA1-driven plasmid replication. Ori, EBNA1-binding origin of replication. EBNA1-binding FR repeats impede the anticlockwise fork. Red triangle indicates the 6×*Ter* array (vertex, non-permissive end). Location of Southern blotting probe shown in black. **b**, Plasmid replication intermediates in 293E cells transiently transfected with 6×*Ter*-containing plasmids or 'no *Ter*' control, co-transfected with empty vector (EV), TusH144A or Tus. Samples from one experiment of DNA digested with XmnI plus SnaBI and analysed by two-dimensional gel electrophoresis/Southern blotting. 6×REV*Ter*: clockwise fork encounters permissive end of *Ter*. Arc A, replication fork. Spot B, Tus/*Ter*-stalled clockwise fork. Spot C, bidirectional fork arrest (double Y) at Tus/6×*Ter*, reflecting incomplete replication block at FR[20]. **c**, Stall spot B quantification, $n = 5$ (see Extended Data Fig. 1a). Error bars represent s.e.m. For Student's *t*-test 6×*Ter* wild-type Tus versus any other, $P < 0.01$; 6×REV*Ter* wild-type Tus versus any other, $P < 0.01$. 6×*Ter* TusH144A versus 6×*Ter* EV, $P < 0.03$. **d**, Upper panel, anti-myc immunoblot of 293E cells expressing empty vector (EV), Tus (WT) or TusH144A (H). Lower panel shows β-tubulin loading control.

[1]Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Boston, Massachusetts 02215, USA. [2]Princeton University, 101 Lewis Thomas Laboratory, Washington Road, Princeton, New Jersey 08544, USA. [3]NIDDK, National Institutes of Health, Building 10, Room 9N105, 10 Center Drive, Bethesda, Maryland 20814, USA. †Present addresses: Temple University, 1801 North Broad Street, Philadelphia, Pennsylvania 19122, USA (G.C.); Brandeis University, 415 South Street, Waltham, Massachusetts 02453, USA (B.H.); University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, Massachusetts 01655, USA (A.K.).

©2014 Macmillan Publishers Limited. All rights reserved

Data Fig. 2). Thus, Tus/*Ter* mediates bidirectional site-specific arrest of mammalian replication forks.

To determine whether Tus/*Ter* induces HR/SCR at a defined chromosomal locus in mammalian cells, we placed 6×*Ter* in a homologous recombination reporter of short- and long-tract gene conversion (termed STGC and LTGC, respectively) between sister chromatids[21]. Duplication of a red fluorescent protein (RFP) cassette distinguishes LTGC (length ≥ 1252 bp; GFP^+RFP^+) from STGC (length <1,252 bp; GFP^+RFP^-; Fig. 2a)[21]. 6×*Ter* abuts an I-SceI site, interrupting an enhanced green fluorescent protein gene (6×*Ter*-I-SceI-*GFP*, Fig. 2a). Recombination of the stalled left-hand fork (Fig. 2a) with the 5′-truncated *GFP* copy (Tr-*GFP*) of the sister chromatid generates wild-type *GFP*. If chromosomal fork arrest were bidirectional, this could produce a two-ended break, generating predominantly STGCs (Fig. 2a and Extended Data Fig. 3a)[21]. In contrast, unidirectional fork arrest with one-ended breaks would favour LTGC, and any STGCs arising from one-ended breaks would necessarily be terminated by non-canonical mechanisms (Extended Data Fig. 3b)[21].

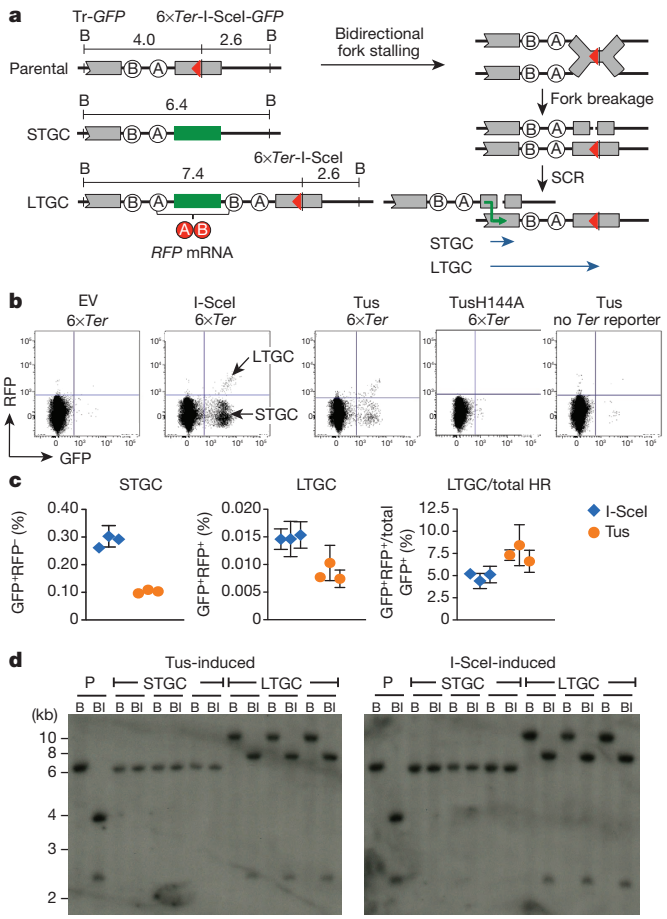


Figure 2 | Tus/*Ter*-induced homologous recombination in mammalian cells. a, 6×*Ter*-HR reporter and major HR products (assuming two-ended breaks). STGC and short-tract gene conversion; LTGC, long-tract gene conversion. LTGC generates wild-type *RFP* expression through RNA splicing. Grey boxes, mutant *GFP*. Green box, wild-type *GFP*. Circles A and B represent 5′ and 3′ artificial *RFP* exons. Tr-*GFP*, 5′ truncated *GFP*. Red triangle indicates 6×*Ter* array adjacent to I-SceI site. B, BglII; *GFP*-hybridizing fragment sizes in kilobases. Bidirectional fork stalling triggers SCR. Green arrow, strand exchange. **b**, FACS data of *Brca1*$^{fl/BRCT}$ 6×*Ter*-HR cells transfected with empty vector (EV), I-SceI, Tus or TusH144A. No *Ter* reporter, *Brca1*$^{fl/BRCT}$ cells carrying the *ROSA26*-targeted HR reporter lacking the *Ter* array[21]. **c**, I-SceI- and Tus-induced HR (blue diamonds and orange circles, respectively) in three independent *Brca1*$^{fl/BRCT}$ clones. Mean of triplicate samples, $n = 3$. Error bars represent s.e.m. Student's *t*-test LTGC/total HR, I-SceI versus Tus: $P = 0.0186$. **d**, Southern blot analysis of Tus- and I-SceI-induced HR in *Brca1*$^{fl/BRCT}$ 6×*Ter*-HR cells (*GFP* probe). P, parental reporter; B, BglII digest; BI, BglII plus I-SceI digest.

We targeted the 6×*Ter*/HR reporter as a single copy to the *ROSA26* locus of mouse embryonic stem (ES) cell line 11CO/47T (*Brca1*$^{fl/BRCT}$)[21]. *Brca1*^{BRCT} encodes a C-terminal truncated protein; the BRCT-encoding elements of *Brca1*^{fl} can be conditionally deleted (generating *Brca1*$^{\Delta}$)[21]. Indeed, Tus, but not TusH144A, induced HR within 6×*Ter*/HR *Brca1*$^{fl/BRCT}$ cells, the major HR product being STGC (Fig. 2b). Tus failed to induce HR in *Brca1*$^{fl/BRCT}$ cells containing a *ROSA26*-targeted HR reporter lacking the *Ter* array (Fig. 2b)[21]. Thus, Tus/*Ter*-induced chromosomal HR requires cognate Tus–*Ter* binding. The ratio LTGC/total HR, a measure of the probability that HR resolves as LTGC, was ~7% in three independent Tus-transfected clones (Fig. 2c and Extended Data Fig. 4a). Three additional independent clones of *Brca1*$^{fl/BRCT}$ ES cells, each containing a single-copy randomly integrated chromosomal 6×*Ter*/HR reporter, behaved similarly (Extended Data Fig. 4b). The predominance of STGC and the consistent results at different loci suggested that Tus/*Ter*-induced HR entails bidirectional fork arrest (Extended Data Fig. 3a). We resolved this definitively by Southern blot analysis of Tus/*Ter*-induced STGCs. Unidirectional fork arrest/breakage (Extended Data Fig. 3b) could produce a one-ended break, generating STGC products of variable size[21]. In contrast, bidirectional fork arrest (Fig. 2a and Extended Data Fig. 3a) could produce a two-ended break, with STGC termination by annealing. This would generate STGC products of fixed size, resembling the parental reporter but lacking the 6×*Ter* array or I-SceI site (Fig. 2a)[21]. Indeed, 44/44 Tus/*Ter*-induced STGCs in 6×*Ter*/HR reporter *Brca1*$^{fl/BRCT}$ cells revealed this latter structure (Fig. 2d). As expected, I-SceI-induced HR behaved similarly (Fig. 2d). A second arrested fork (right-hand fork, Fig. 2a) must provide the homologous second end during Tus/*Ter*-induced STGC. Therefore, Tus/*Ter*-induced STGC is the product of bidirectional replication fork arrest. Overall, I-SceI-induced HR in *Brca1*$^{fl/BRCT}$ 6×*Ter*/HR reporter cells was ~20% of that in isogenic *ROSA26*-targeted *Brca1*$^{fl/BRCT}$ HR reporter cells, which lack a 6×*Ter* array (data not shown)[21]. The reasons for this difference are unclear.

To investigate further the non-polar behaviour of Tus/*Ter* in mammalian HR, we studied the Tus mutant F140A that binds duplex *Ter* with higher affinity than Tus but is defective for the *Ter* C-6 base-flipping 'locking' mechanism that contributes to polar fork arrest in *E. coli*[16]. TusF140A induced higher levels of HR than Tus in 6×*Ter*/HR *Brca1*$^{fl/BRCT}$ cells (Extended Data Fig. 4c, d), showing that the C-6 'lock' is dispensable for Tus/*Ter*-induced HR in mammalian cells. This might be explained by the different polarities of the *E. coli* DnaB and vertebrate MCM replicative helicases. *Ter* C-6 is located on the leading strand of the fork approaching the non-permissive end of *Ter*. Unlike DnaB, which translocates along the lagging strand, the MCM helicase translocates along the leading strand and might occlude *Ter* C-6 within its barrel, thereby denying Tus access to the C-6 lock mechanism. To determine the minimal number of Tus/*Ter* complexes needed for HR induction, we generated reporters containing 3, 2 or 1 *Ter* sites (Extended Data Fig. 4e). We targeted each, in parallel, as a single copy to the *ROSA26* locus of *Brca1*$^{fl/BRCT}$ ES cells and found that a minimum of 3 *Ter* sites is required for robust Tus-induced HR (Extended Data Fig. 4f). We inverted the 6×*Ter* array orientation to generate a 6×REV*Ter*/HR reporter. When targeted as a single copy to the *ROSA26* locus of *Brca1*$^{fl/BRCT}$ ES cells, this reporter supported Tus-induced HR as robustly as the 6×*Ter*/HR reporter (Extended Data Fig. 4f). Our findings do not exclude a polar component to Tus/*Ter*-induced fork stalling on a mammalian chromosome but this polarity, if present, is relative and not absolute.

Interstrand DNA crosslink repair of plasmids replicating in *Xenopus laevis* egg extracts entails endonucleolytic attack of bidirectionally stalled forks[22]. Interestingly, Tus/*Ter*-induced HR was suppressed by depletion of the endonuclease scaffold Slx4/FancP[23,24] to a greater extent than I-SceI-induced HR (Extended Data Fig. 5), indicating that Slx4 contributes specifically to Tus/*Ter*-induced HR. However, it is not clear whether Slx4 mediates endonucleolytic attack of stalled forks during

©2014 Macmillan Publishers Limited. All rights reserved

Tus/*Ter*-induced HR. Work in *Schizosaccharomyces pombe* suggests that alternative mechanisms, such as template switching, could mediate HR at stalled mammalian forks[7].

To determine whether BRCA1 regulates HR at stalled replication forks, we transduced 6×*Ter*/HR *Brca1*$^{fl/BRCT}$ cells with adeno-Cre and screened for *Brca1* loss (Fig. 3a). The resulting *Brca1*$^{\Delta/BRCT}$ cells are viable hypomorphs with growth characteristics similar to *Brca1*$^{fl/BRCT}$ cells[21]. We studied Tus-induced HR in three independent Cre-treated clones of each genotype (Fig. 3). Surprisingly, Tus-induced STGC in 6×*Ter*/HR *Brca1*$^{\Delta/BRCT}$ cells showed no reduction compared to *Brca1*$^{fl/BRCT}$ cells, but LTGC was elevated twofold (Fig. 3b). Correspondingly, the probability of engaging LTGC during Tus/*Ter*-induced HR was doubled to ~15% (Fig. 3b). Consistent with our recent findings, I-SceI-induced HR in 6×*Ter*/HR *Brca1*$^{\Delta/BRCT}$ cells was diminished and biased in favour of LTGC (Fig. 3b)[21]. Southern blot analysis of Tus/*Ter*-induced STGC and LTGC products in 6×*Ter*/HR *Brca1*$^{\Delta/BRCT}$ cells revealed patterns similar to *Brca1*$^{fl/BRCT}$ cells (Extended Data Fig. 6). However, in *Brca1*$^{\Delta/BRCT}$ cells, 6/41 (15%) Tus/*Ter*-induced STGC and 3/15 (20%) LTGC clones retained an additional copy of the parental reporter (Extended Data Fig. 6). This was not separable by recloning, suggesting that it was retained by non-disjunction. A total of 4/41 (9.8%) I-SceI-induced STGC *Brca1*$^{\Delta/BRCT}$ clones revealed non-disjunction; thus, non-disjunction is not specific to Tus/*Ter*-induced HR. The fact that the donor sister was unaltered during LTGC excludes crossing-over as a cause of the LTGC outcome in these clones[25,26].

We asked whether Brca1 domains additional to the BRCT repeat regulate Tus/*Ter*-induced HR. Indeed, short interfering (siRNA)-mediated Brca1 depletion suppressed STGC but increased LTGC in both *Brca1*$^{fl/BRCT}$ and *Brca1*$^{\Delta/BRCT}$ cells (Extended Data Fig. 7). In Brca1-depleted *Brca1*$^{\Delta/BRCT}$ cells, ~40% of all HR products were LTGCs. More than half of the BRCA1 polypeptide is encoded by exon 11, which is a target of inactivating germline mutations in hereditary breast/ovarian cancer; exon 11 is also alternatively spliced, generating an in-frame nuclear Δexon11 gene product that retains an N-terminal RING domain and C-terminal BRCT functions[27]. To test whether *Brca1* exon 11 regulates Tus/*Ter*-induced HR, we targeted a single copy of the 6×*Ter*/HR reporter to the *ROSA26* locus of mouse *Brca1*$^{fl/Exon11}$ ES cells. The *Brca1*^{Exon11} allele lacks exon 11; exon 11 of *Brca1*^{fl} can be conditionally deleted to generate *Brca1*$^{\Delta}$ (Fig. 4a)[28]. Note that *Brca1*^{fl} and *Brca1*$^{\Delta}$ denote distinct *Brca1* alleles in the two *Brca1* conditional systems described here. Following adeno-Cre treatment, we retrieved 6×*Ter*/HR *Brca1*$^{fl/Exon11}$ and 6×*Ter*/HR *Brca1*$^{\Delta/Exon11}$ clones. Each of three independent 6×*Ter*/HR *Brca1*$^{\Delta/Exon11}$ clones revealed reduced Tus/*Ter*-induced STGC but increased absolute frequencies of LTGC in comparison with three independent 6×*Ter*/HR *Brca1*$^{fl/Exon11}$ clones (Fig. 4b). Deletion of *Brca1* exon 11 increased the probability of engaging Tus/*Ter*-induced LTGC ~4-fold to ~30% (Fig. 4b). In contrast, the absolute frequency of I-SceI-induced LTGC was reduced in *Brca1*$^{\Delta/Exon11}$ cells and ~20% of HR products were

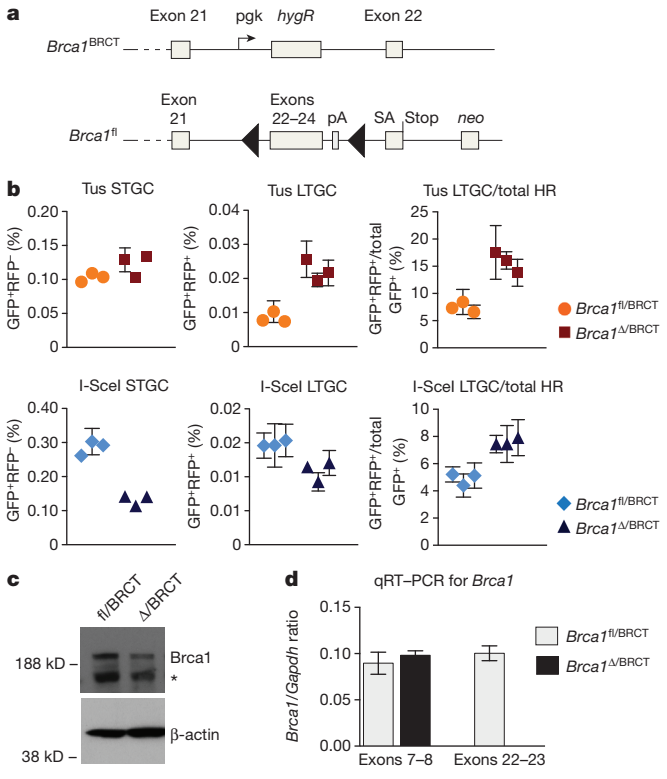


Figure 3 | The Brca1 tandem BRCT repeat regulates Tus/*Ter*-induced homologous recombination. **a**, *Brca1* gene in *Brca1*$^{fl/BRCT}$ ES cells. *Brca1*^{BRCT} encodes a truncated protein. Cre converts *Brca*^{fl} to the exon 22–24-deleted *Brca1*$^{\Delta}$ allele. Grey boxes, *Brca1* exons; black triangles, *loxP* sites; pA, polyadenylation signal; SA, splice acceptor; *neo:* neomycin resistance gene; pgk, phosphoglycerate kinase promoter. **b**, Tus- and I-SceI-induced HR in *Brca1*$^{fl/BRCT}$ and *Brca1*$^{\Delta/BRCT}$ 6×*Ter*-HR cells (three independent clones each). Mean of triplicate samples, $n = 4$. Error bars represent s.e.m. Student's *t*-test *Brca1*$^{fl/BRCT}$ versus *Brca1*$^{\Delta/BRCT}$ in all 6 panels $P < 0.05$. **c**, Upper panel, endogenous Brca1 immunoblot in *Brca1*$^{fl/BRCT}$ and *Brca1*$^{\Delta/BRCT}$ ES cells. Asterisk indicates a background band. Lower panel, β-actin loading control. **d**, Quantitative polymerase chain reaction with reverse transcription (qRT–PCR) for *Brca1* mRNA. Exon 22–23 is deleted in *Brca1*$^{\Delta/BRCT}$ cells.

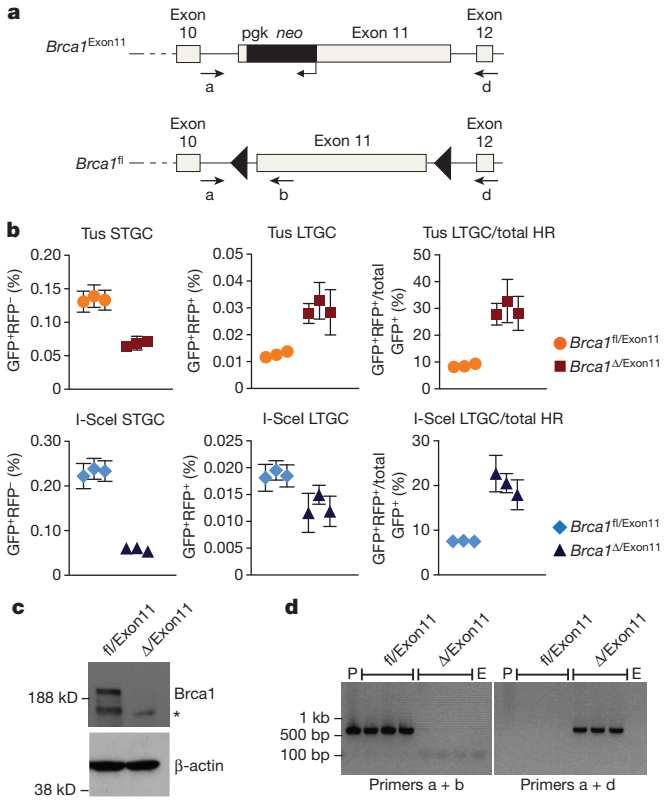


Figure 4 | *Brca1* Exon11 regulates Tus/*Ter*-induced homologous recombination. **a**, *Brca1* gene in *Brca1*$^{fl/Exon11}$ ES cells. The *Brca1*^{Exon11} encodes the Δexon11 product. Cre converts *Brca1*^{fl} to exon11-deleted *Brca1*$^{\Delta}$ allele. Symbols as in Fig. 3. PCR primers a, b and d shown. **b**, Tus- and I-SceI-induced HR in *Brca1*$^{fl/Exon11}$ and *Brca1*$^{\Delta/Exon11}$ 6×*Ter*-HR cells (three independent clones each). Mean of triplicate samples, $n = 4$. Error bars represent s.e.m. Student's *t*-test *Brca1*$^{fl/Exon11}$ versus *Brca1*$^{\Delta/Exon11}$, in all 6 panels $P < 0.005$. **c**, Upper panel, endogenous Brca1 immunoblot in *Brca1*$^{fl/Exon11}$ and *Brca1*$^{\Delta/Exon11}$ ES cells. The asterisk denotes a background band. The lower panel denotes a β-actin loading control. **d**, PCR genotyping of *Brca1*$^{fl/Exon11}$ and *Brca1*$^{\Delta/Exon11}$ clones from panel b. P, untargeted *Brca1*$^{fl/Exon11}$. E, empty (no DNA) control. *Brca1*^{fl} product, 531 bp; *Brca1*$^{\Delta}$ product, 621 bp.

©2014 Macmillan Publishers Limited. All rights reserved

LTGCs (Fig. 4b). Thus, *Brca1* exon 11 contributes to Tus/*Ter*-induced HR both quantitatively and qualitatively.

To determine whether BRCA2/Rad51 regulates Tus/*Ter*-induced HR, we used siRNA to deplete Brca2 or Rad51 during HR induction. Depletion of Brca2 suppressed Tus/*Ter*-induced STGC but elevated LTGC frequencies in both $Brca1^{fl/BRCT}$ and $Brca1^{\Delta/BRCT}$ cells (Extended Data Fig. 8). In $Brca1^{fl/BRCT}$ and $Brca1^{\Delta/BRCT}$ cells depleted of Brca2, ~30% and ~50% respectively of all Tus/*Ter*-induced HR products were LTGCs, whereas the equivalent probabilities for Rad51-depleted cells were ~40% and ~70% (Extended Data Fig. 9). Thus, suppression of LTGC at stalled forks is a shared function of BRCA1, BRCA2 and Rad51. Inhibition of 53BP1 partially reversed defective I-SceI-induced HR in $Brca1^{\Delta/BRCT}$ cells, as expected[21], but did not affect Tus/*Ter*-induced HR in either $Brca1^{fl/BRCT}$ or $Brca1^{\Delta/BRCT}$ cells (Extended Data Fig. 10). This suggests that BRCA1's functions in Tus/*Ter*-induced and SceI-induced HR are, in part, distinct. LTGC at stalled forks may include pathological responses analogous to break-induced replication in yeast[3,8,29,30]. Our work identifies loss of BRCA1/BRCA2/Rad51-dependent suppression of LTGC at stalled replication forks as a potential contributor to breast/ovarian cancer predisposition.

METHODS SUMMARY

A description of the molecular biology methods, cell culture and recombination assays can be found in the full Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 July 2013; accepted 31 March 2014.
Published online 28 April 2014.

1. Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol. Cell* **40,** 179–204 (2010).
2. Nagaraju, G. & Scully, R. Minding the gap: the underground functions of BRCA1 and BRCA2 at stalled replication forks. *DNA Repair* **6,** 1018–1031 (2007).
3. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5,** e1000327 (2009).
4. Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25,** 156–165 (2000).
5. Cox, M. M. *et al.* The importance of repairing stalled replication forks. *Nature* **404,** 37–41 (2000).
6. Michel, B., Grompone, G., Flores, M. J. & Bidnenko, V. Multiple pathways process stalled replication forks. *Proc. Natl Acad. Sci. USA* **101,** 12783–12788 (2004).
7. Lambert, S. *et al.* Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Mol. Cell* **39,** 346–359 (2010).
8. Carr, A. M. & Lambert, S. Replication stress-induced genome instability: the dark side of replication maintenance by homologous recombination. *J. Mol. Biol.* **425,** 4733–4744 (2013).
9. Scully, R. & Livingston, D. M. In search of the tumour-suppressor functions of BRCA1 and BRCA2. *Nature* **408,** 429–432 (2000).
10. Lomonosov, M., Anand, S., Sangrithi, M., Davies, R. & Venkitaraman, A. R. Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. *Genes Dev.* **17,** 3017–3022 (2003).
11. Pathania, S. *et al.* BRCA1 is required for postreplication repair after UV-induced DNA damage. *Mol. Cell* **44,** 235–251 (2011).
12. Schlacher, K. *et al.* Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145,** 529–542 (2011).
13. Amin, A. A. & Hurwitz, J. Polar arrest of the simian virus 40 tumor antigen-mediated replication fork movement *in vitro* by the tus protein-*terB* complex of *Escherichia coli*. *J. Biol. Chem.* **267,** 18612–18622 (1992).
14. Hill, T. M. & Marians, K. J. *Escherichia coli* Tus protein acts to arrest the progression of DNA replication forks *in vitro*. *Proc. Natl Acad. Sci. USA* **87,** 2481–2485 (1990).
15. Bidnenko, V., Ehrlich, S. D. & Michel, B. Replication fork collapse at replication terminator sequences. *EMBO J.* **21,** 3898–3907 (2002).
16. Mulcair, M. D. *et al.* A molecular mousetrap determines polarity of termination of DNA replication in *E. coli*. *Cell* **125,** 1309–1319 (2006).
17. Dhar, S. K. *et al.* Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin. *Cell* **106,** 287–296 (2001).
18. Follonier, C. & Lopes, M. Combined bidimensional electrophoresis and electron microscopy to study specific plasmid DNA replication intermediates in human cells. *Methods Mol. Biol.* **1094,** 209–219 (2014).
19. Kirchmaier, A. L. & Sugden, B. Plasmid maintenance of derivatives of oriP of Epstein-Barr virus. *J. Virol.* **69,** 1280–1283 (1995).
20. Gahn, T. A. & Schildkraut, C. L. The Epstein-Barr virus origin of plasmid replication, oriP, contains both the initiation and termination sites of DNA replication. *Cell* **58,** 527–535 (1989).
21. Chandramouly, G. *et al.* BRCA1 and CtIP suppress long-tract gene conversion between sister chromatids. *Nature Commun.* **4,** 2404 (2013).
22. Räschle, M. *et al.* Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134,** 969–980 (2008).
23. Castor, D. *et al.* Cooperative control of Holliday junction resolution and DNA repair by the SLX1 and MUS81–EME1 nucleases. *Mol. Cell* **52,** 221–233 (2013).
24. Garner, E., Kim, Y., Lach, F. P., Kottemann, M. C. & Smogorzewska, A. Human GEN1 and the SLX4-associated nucleases MUS81 and SLX1 are essential for the resolution of replication-induced Holliday junctions. *Cell Rep.* **5,** 207–215 (2013).
25. Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19,** 3398–3407 (2000).
26. Nagaraju, G., Odate, S., Xie, A. & Scully, R. Differential regulation of short- and long-tract gene conversion between sister chromatids by Rad51C. *Mol. Cell. Biol.* **26,** 8075–8086 (2006).
27. Huber, L. J. *et al.* Impaired DNA damage response in cells expressing an exon 11-deleted murine Brca1 variant that localizes to nuclear foci. *Mol. Cell. Biol.* **21,** 4005–4015 (2001).
28. Xu, X. *et al.* Conditional mutation of *Brca1* in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. *Nature Genet.* **22,** 37–43 (1999).
29. Llorente, B., Smith, C. E. & Symington, L. S. Break-induced replication: what is it and what is it for? *Cell Cycle* **7,** 859–864 (2008).
30. Saini, N. *et al.* Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502,** 389–392 (2013).

Acknowledgements We thank B. Michel for reagents and advice; D. Livingston, K. Marians, J. Walter, I. Hickson, S. Powell, V. Zakian and members of the Scully laboratory for helpful discussions; I. Hickson and S. Powell for sharing their unpublished data on Tus/*Ter* replication blocks in eukaryotes; R. Baer for antibodies and A. Ashworth for *Brca1* conditional ES cell line 11CO/47T. This work was supported by NIH grants R01CA095175, R01GM073894 and R21CA144017 (to R.S.). N.A.W. was supported by an NIH/NCI postdoctoral fellowship (5T32CA081156) and an ACS postdoctoral research fellowship (PF-12-248-01-DMC). C.F. was supported by NIH grant R37GM26938 (to V.A.Z.).

Author Contributions N.A.W., G.C. and R.S. designed experiments. N.A.W., G.C., B.H. and A.K. performed experiments. C.F. provided expert advice on execution of two-dimensional gel electrophoresis experiments. C.D. generated *Brca1* exon 11 conditional ES cells. N.A.W. analysed the data. N.A.W. and R.S wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S. (rscully@bidmc.harvard.edu).

©2014 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature13409

Structural rearrangements of a polyketide synthase module during its catalytic cycle

Jonathan R. Whicher[1,2]*, Somnath Dutta[1]*, Douglas A. Hansen[1,3], Wendi A. Hale[4], Joseph A. Chemler[1], Annie M. Dosey[1], Alison R. H. Narayan[1], Kristina Håkansson[4], David H. Sherman[1,3,4,5], Janet L. Smith[1,6] & Georgios Skiniotis[1,6]

The polyketide synthase (PKS) mega-enzyme assembly line uses a modular architecture to synthesize diverse and bioactive natural products that often constitute the core structures or complete chemical entities for many clinically approved therapeutic agents[1]. The architecture of a full-length PKS module from the pikromycin pathway of *Streptomyces venezuelae* creates a reaction chamber for the intramodule acyl carrier protein (ACP) domain that carries building blocks and intermediates between acyltransferase, ketosynthase and ketoreductase active sites (see accompanying paper[2]). Here we determine electron cryo-microscopy structures of a full-length pikromycin PKS module in three key biochemical states of its catalytic cycle. Each biochemical state was confirmed by bottom-up liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry. The ACP domain is differentially and precisely positioned after polyketide chain substrate loading on the active site of the ketosynthase, after extension to the β-keto intermediate, and after β-hydroxy product generation. The structures reveal the ACP dynamics for sequential interactions with catalytic domains within the reaction chamber, and for transferring the elongated and processed polyketide substrate to the next module in the PKS pathway. During the enzymatic cycle the ketoreductase domain undergoes dramatic conformational rearrangements that enable optimal positioning for reductive processing of the ACP-bound polyketide chain elongation intermediate. These findings have crucial implications for the design of functional PKS modules, and for the engineering of pathways to generate pharmacologically relevant molecules.

Modular PKSs produce chemically diverse polyketide natural products that account for a large number of pharmaceutical compounds[1]. The modular organization of these enzyme assembly lines provides important opportunities for synthetic biology applications and bioengineering[3–5].

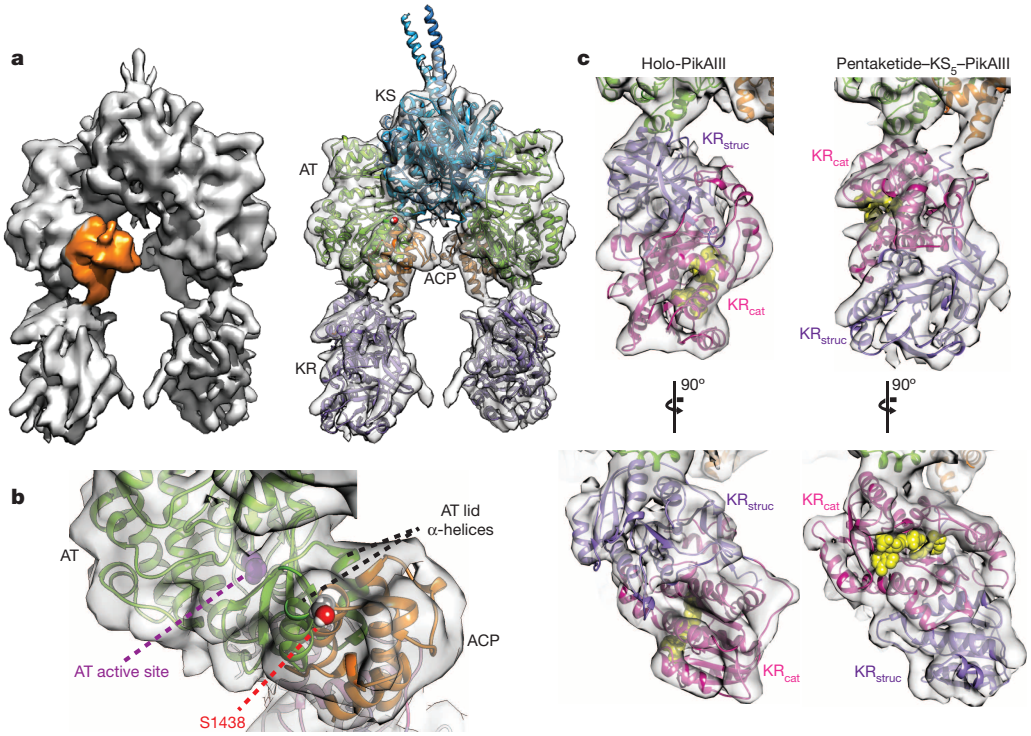


Figure 1 | Cryo-EM structure of pentaketide–KS_5–PikAIII. a, Solid rendering (left) and transparent representation with modelled structures (right) of the pentaketide–KS_5–PikAIII cryo-EM map. **b**, Interface of AT[12] and ACP[14] docked within the cryo-EM density. ACP helix 2 (residues 1439–1452) contacts the two AT lid helices (residues 701–709 and 732–744). The phosphopantetheinylated Ser 1438 of ACP is 35 Å from the AT active site. **c**, Comparison of the KR structure[13] docked in the cryo-EM map for holo-PikAIII and pentaketide–KS_5–PikAIII. The dramatic rearrangement upon pentaketide–KS formation brings two helices (residues 1126–1138, 1156–1166) of the KR catalytic subdomain (KR_{cat}, magenta) in contact with an AT helix (residues 760–775) and moves the structural subdomain of KR (KR_{struc}, purple) away from AT.

[1]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109, USA. [2]Chemical Biology Graduate Program, University of Michigan, Ann Arbor, Michigan 48109, USA. [3]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [4]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [5]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan 48109, USA. [6]Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA.
*These authors contributed equally to this work.

©2014 Macmillan Publishers Limited. All rights reserved

However, many combinatorial biosynthesis efforts have been beset by problems of inefficiency and a limited ability to display 'plug and play' behaviour[6–8]. Thus, there is a critical need to dissect and understand the bacterial type I PKS module architecture and its conformational dynamics during polyketide chain elongation and processing.

Each module in the bacterial type I PKS catalyses the elongation of an intermediate by two carbon atoms in a linear sequence, and may also modify the intermediate by reductive processing or β-branching[9]. An ACP domain within each module shuttles extender units and polyketide chain elongation intermediates between the catalytic domains. In an accompanying paper[2], we describe the overall architecture of PKS module 5 (PikAIII) from the pikromycin pathway (Pik)[10]. This PKS module is organized around a dimeric ketosynthase (KS) domain and a central chamber for the ACP, with the active sites of catalytic domains facing the chamber. Here we examined the dynamics of the PikAIII module 5 ACP (ACP_5) within the reaction chamber of PikAIII to probe its localization and underlying protein–protein interactions during the catalytic cycle. Within PikAIII, the acyltransferase (AT_5) loads ACP_5 with a methylmalonyl (MM) extender unit, and the KS (KS_5) catalyses a decarboxylative condensation reaction between the MM group and the pentaketide delivered by the module 4 ACP to produce a β-ketohexaketide intermediate. The β-hydroxyhexaketide product is subsequently generated through NADPH-dependent reduction by the ketoreductase (KR_5) domain of PikAIII.

Incubation of full-length holo-PikAIII with thiophenol-activated pentaketide[11] generated pentaketide–PikAIII, in which this natural substrate was bound to the catalytic Cys 209 of KS_5 (Extended Data Fig. 1a–c). The three-dimensional electron cryo-microscopy (cryo-EM) map of pentaketide–KS_5–PikAIII was calculated at a resolution of 7.9 Å (Extended Data Fig. 2a, d), enabling precise docking of the structures of 6-deoxyerythronolide B synthase (DEBS) module 5 KS and AT[12], DEBS module 1 KR[13], and DEBS module 2 ACP[14]. This EM reconstruction revealed several striking conformational changes compared to the structure of holo-PikAIII (Fig. 1a)[2]. The AT_5 moves towards KS_5 by ~2 Å and partially occludes the side entrance to the KS_5 active site (Extended Data Fig. 3a), which may sequester the labile polyketide intermediate until the extension reaction occurs. In addition, several loops on the KS surface appear to have changed position and additional densities appear at the bottom of the KS, below the catalytic Cys 209 (Extended Data Fig. 3a, b). These KS conformational changes upon pentaketide loading on Cys 209 may be allosterically sensed by the AT domains, leading to the AT_5 shift. Pentaketide-induced conformational changes are also evident in the position of ACP_5, which shifts 10 Å and 30 Å towards AT_5 from its positions in the two holo-PikAIII conformers[2]. In this position ACP_5 appears to insert helix 2 between the two helices of the AT_5 lid subdomain (Fig. 1b). The interaction requires small conformational changes to one or both domains, although the details are not discernable in the EM maps. The reactive Ser 1438 of ACP_5 is directed towards the AT_5 active site, presumably poising the ACP for loading with the MM extender unit by the AT domain. We probed the ACP–AT interface with single alanine substitutions in AT_5 of the PikAIII–TE protein used for functional assays[15,16] (PikAIII fused to PikAIV TE, the module 6 thioesterase domain). AT_5 Glu 735 is in the ACP_5 contact zone, while Arg 747, approximately 20 Å from the AT_5 active site, was predicted to reside in the ACP docking site of a DEBS AT[17]. Consistent with the cryo-EM results, the E735A or the R747A PikAIII–TE mutants exhibited tenfold and fivefold reduced formation of the 10-deoxymethynolide (10-dml) macrolactone product, respectively (Extended Data Fig. 4).

The most surprising feature of the pentaketide–KS_5–PikAIII structure is that KR_5 has undergone an end-to-end flip compared to the holo-PikAIII reconstruction (Fig. 1c). This finding suggests that the pentaketide-induced repositioning of AT_5 and ACP_5 also affects the AT_5–KR_5 linker, causing the large rotation of KR_5. The end-to-end flip does not tangle the PikAIII topology, as the linker connections (residues 916 and 1360) are adjacent in KR_5 and reside at the junction of the structural (KR_{5struc})

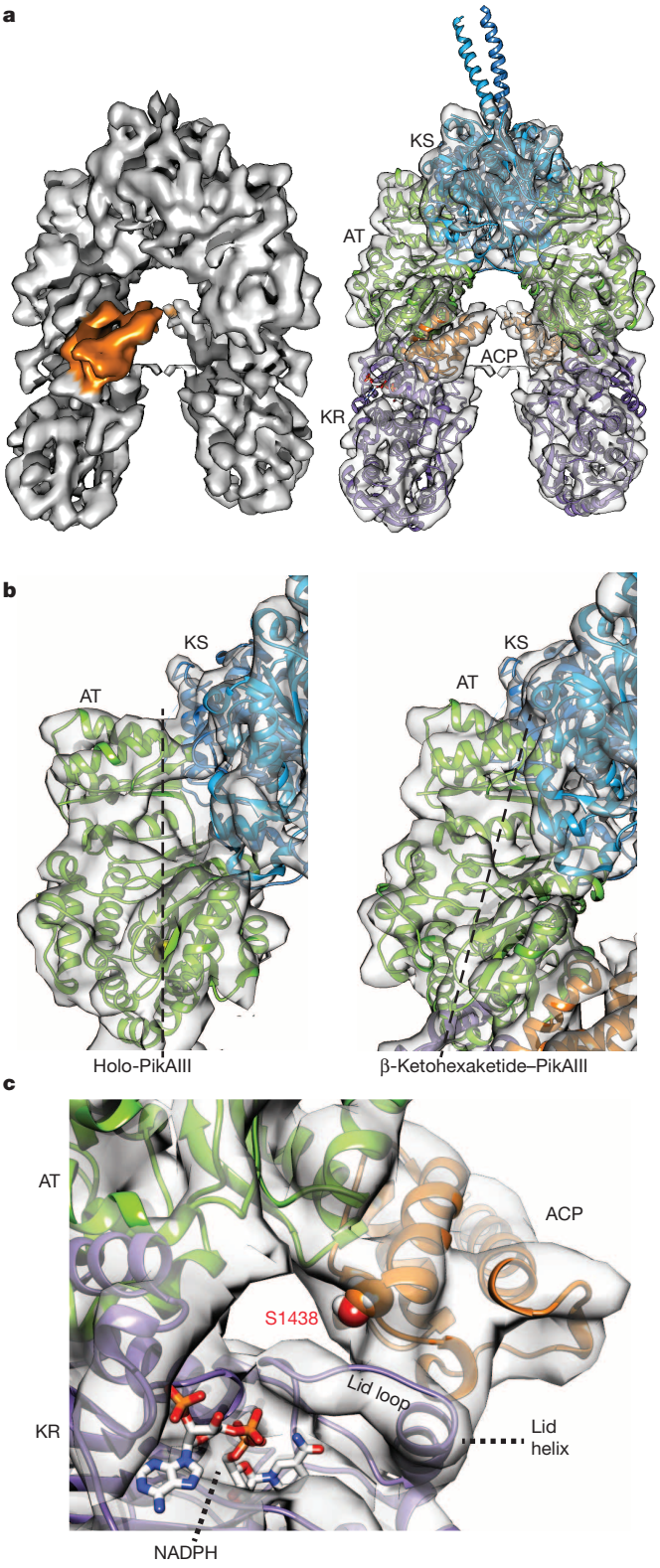


Figure 2 | Cryo-EM structure of β-ketohexaketide–PikAIII. **a**, Solid rendering (left) and transparent representation (right) with modelled structures of the β-ketohexaketide–PikAIII cryo-EM map. **b**, Comparison of the KS–AT interface between holo-PikAIII (left) and β-ketohexaketide–PikAIII (right). Polyketide chain extension results in an 8 Å shift and rotation of AT towards the KS. **c**, ACP interaction with KR. ACP interacts with the KR lid helix (residues 1302–1309) and lid loop (1293–1301) to position ACP Ser 1438 near the NADPH-binding cleft.

©2014 Macmillan Publishers Limited. All rights reserved

and catalytic (KR_{5cat}) subdomains. The ~180° KR_5 rotation positions the catalytic domain proximal to AT_5, with the AT_5–KR_5 interface formed by an AT_5 helix and two KR_{5cat} helices. Specifically, KR_5 Arg 1133 and His 1137 are positioned to interact with AT_5 Glu 766 or Glu 768. We confirmed the importance of the KR_{5cat}–AT_5 interface for catalytic function with R1133E, H1137E, E766R and E768R substitutions, which had 2- to 15-fold reduced activity relative to the wild-type PikAIII–TE (Extended Data Fig. 4). By contrast, substitution of a KR_{5struc} residue at the KR_{5struc}–AT_5 interface of holo-PikAIII had no impact on production of 10-dml[2]. The end-to-end KR_5 flip, together with the unrelated repositioning of KR_5 observed upon ACP_5 deletion[2], indicates that its orientation depends on the state of other PikAIII domains.

To characterize the architecture of PikAIII, in which the β-ketohexaketide chain elongation intermediate is tethered to ACP_5, we incubated holo-PikAIII with both MM–coenzyme A (CoA) and thiophenol-activated pentaketide (Extended Data Fig. 1d–i). The three-dimensional cryo-EM structure of β-ketohexaketide–PikAIII at 7.8 Å resolution (Extended Data Fig. 2b, d) reveals the same overall PikAIII conformation of the catalytic domains as observed for pentaketide–KS_5–PikAIII, with the KR_5 catalytic domain proximal to AT_5 (Fig. 2a). Furthermore, in the β-ketohexaketide–PikAIII structure, AT_5 is shifted towards KS_5 by 8 Å compared with the holo state and occludes the KS_5 side entrance (Fig. 2b), probably preventing intermediate transfer from the upstream PikAII ACP_4 to the KS_5 active site.

KR domains catalyse the NADPH-dependent reduction of the β-keto group to form either a D- (B-type KR) or L- (A-type KR) hydroxyl group. To form the two hydroxyl group stereochemistries, A- and B-type KRs are proposed to guide substrates into their active sites from opposite directions[18,19] through the differential ordering of the KR 'lid loop' and 'lid helix' upon cofactor and substrate binding[20]. The β-ketohexaketide ACP_5 docks with the 'lid helix' and 'lid loop' of the A-type KR_5, with ACP_5 Ser 1438 pointing towards the KR_5 active site at a distance of 19 Å (Fig. 2c). Given the long linker between KR and ACP (43 amino acids), it is unclear whether ACP_5 interacts with the KR_5 of the same or the opposite monomer. An R1308E substitution in the KR_5 lid helix reduced activity tenfold compared with wild type, validating the observed docking site (Extended Data Fig. 4a, b). From this position, a β-keto ACP substrate can enter the KR active site from either direction, facilitating formation of β-hydroxy products of KR-specified chirality[19]. Thus, the observed ACP docking site on the KR is a general solution for PKS modules with either A- or B-type KRs.

To form β-hydroxyhexaketide–PikAIII, we incubated holo-PikAIII with MM–CoA, thiophenol-activated pentaketide and NADPH (Extended Data Fig. 1j), and applied cryo-EM to obtain its structure. During initial three-dimensional reconstructions we observed three different positions for ACP_5 and applied multiple-reference-supervised classification to calculate three independent three-dimensional reconstructions at resolutions of ~11 Å (Fig. 3a and Extended Data Fig. 2c, d)[21,22]. In all three conformers the catalytic domains are identically positioned, with the KS_5 side entrance occluded and the KR_5 domain oriented with its active site proximal to AT_5, as is the case with the pentaketide–PikAIII and β-ketohexaketide–PikAIII structures. In each conformer the ACP_5 domains are below the KR_{5struc} and completely outside the catalytic chamber (Fig. 3b). The ACP structure can be precisely docked within each corresponding density in all three states, indicating that the three positions are highly populated and that ACP_5 is not particularly flexible under these conditions (Fig. 3b, c). In these configurations, Ser 1438 of ACP_5 points away from PikAIII, which is fully compatible with ACP_5 docking

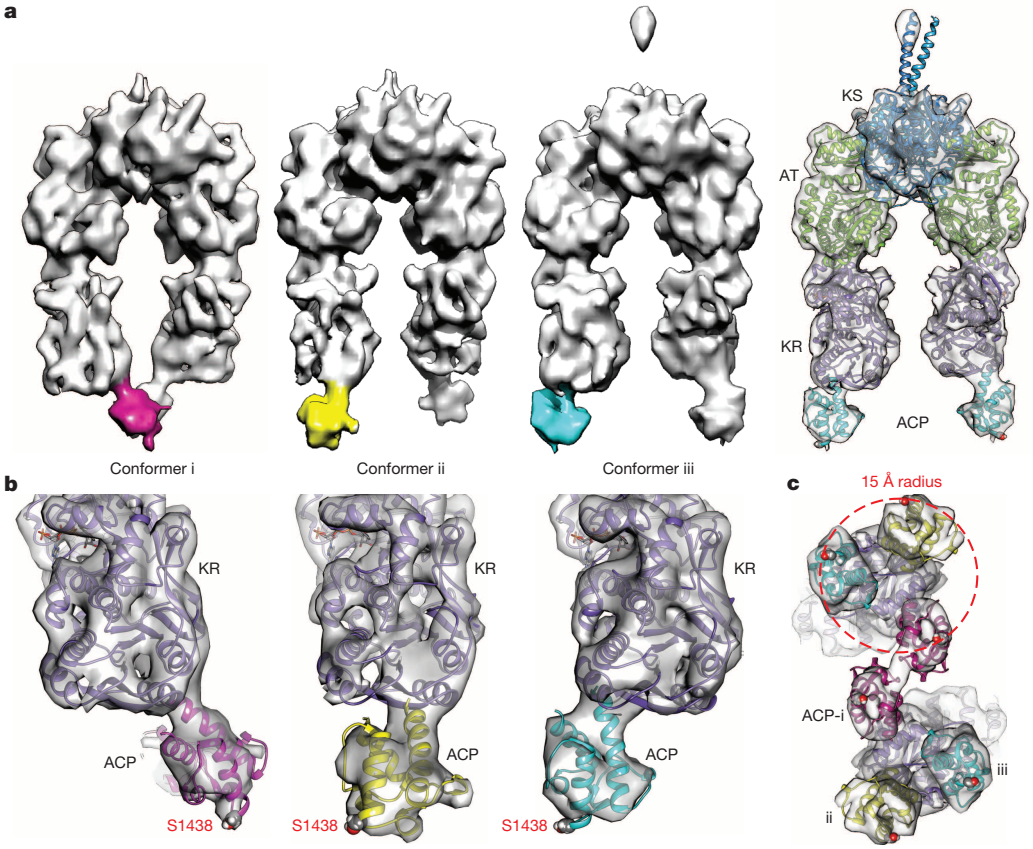


Figure 3 | Cryo-EM structures of β-hydroxyhexaketide–PikAIII. **a**, Solid rendering of cryo-EM maps for the three conformers (i, ii and iii) of β-hydroxyhexaketide–PikAIII. The ACP density is differentially coloured with magenta, yellow and cyan. Right, a transparent representation with modelled structures of the adjacent β-hydroxyhexaketide–PikAIII conformer. **b**, Docking of the ACP structure[14] in the corresponding density map of each β-hydroxyhexaketide–PikAIII conformer shows that Ser 1438 invariably points away from the module. The ACP structures are coloured according to the conformers in **a**. **c**, Bottom view of overlay of the three cryo-EM maps of β-hydroxyhexaketide–PikAIII. The three different ACP positions cover a surface with a radius of 15 Å.

©2014 Macmillan Publishers Limited. All rights reserved

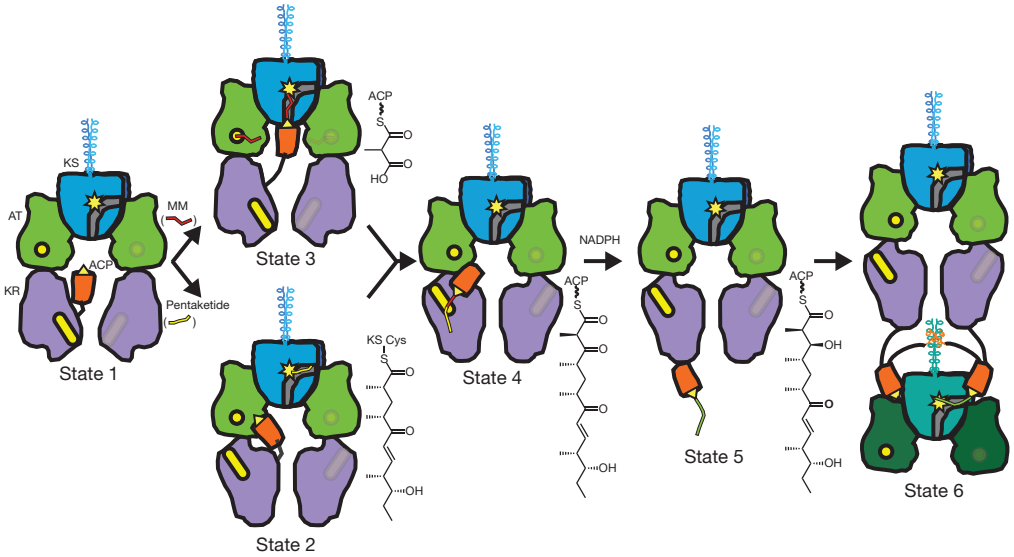


Figure 4 | Schematic of substrate processing in PikAIII. Transfer of the upstream intermediate (pentaketide, yellow line) to the KS domain (blue with yellow active site) of holo-PikAIII (state 1) induces an end-to-end flip of the KR domain (purple, yellow active site) and positions the ACP domain (orange, yellow active serine) at the AT domain (green, yellow active site) (state 2). MM (red line) loading positions the ACP at the bottom entrance of the KS (state 3). KS-catalysed extension of the pentaketide to a β-ketohexaketide (red–yellow line) positions the ACP at the KR domain (state 4). KR-catalysed reduction of the β-ketohexaketide to β-hydroxyhexaketide (green line) ejects the ACP out of the central chamber (state 5) and into position for intermediate transfer to PikAIV (state 6).

to the KS_6 side entrance of PikAIV for substrate transfer[2]. This demonstrates the mechanism to maintain the directionality of PKS pathways, because only when the substrate is completely modified is the ACP ejected from the reaction chamber and positioned for transfer to the next module (Fig. 4). Consistent with this mechanism, when NADPH was excluded from a coupled PikAIII/PikAIV assay[15], thus preventing β-hydroxyhexaketide formation, substrate consumption was halved, indicating that intermediate transfer from PikAIII to PikAIV is impaired when ACP is sequestered within the reaction chamber (Extended Data Fig. 5).

Our studies capture the travels of PikAIII ACP_5 in the reaction chamber of a complete PKS module and reveal a stunning ACP localization dependent on the identity of the phosphopantetheine-linked polyketide intermediate or extender unit (Fig. 4). The cryo-EM structures identified interaction surfaces for ACP_5 on each of the catalytic domains, differing in all cases from predicted sites, and confirmed by site-directed mutagenesis. The interacting surfaces of ACP_5 are generally charge-complementary with partner catalytic domains, and similar to those recently observed in other ACP–enzyme interactions[2,23–25]. Interestingly, ACP_5 does not engage in specific protein–protein interactions with catalytic domains unless the appropriate substrate is loaded, indicating that tethered substrate is a primary determinant of ACP_5 localization. However, the high level of ordering observed for ACP does not apply to the post-ACP elements. The post-ACP helices, while essential for PikAIII dimerization, do not appear to be in a fixed position relative to the other PikAIII domains, suggesting that docked or fused downstream modules would also not assume a fixed position relative to the upstream module. Remarkably, the cryo-EM structures also demonstrate the existence of interdomain cross-talk within the module, evidenced both by the end-to-end flip of the KR_5 that positions its active site to interact with the β-ketohexaketide-loaded ACP and by movements of AT_5 that appear to limit access to the KS_5 active-site side entrance. Both structural transitions occur when the module contains pentaketide or hexaketide intermediates, indicating that the operation of the assembly line sets the conformational state of the PKS module and positions the domains for the next catalytic step. This clever design of the PKS module leads to optimal throughput, in which substrate availability and the rates of catalysis, rather than transfer rates, are probably the limiting steps in the catalytic cycle of type I modular megasynthases. The striking influence of substrate in the positioning and orientation of ACP, KR and AT has profound implications for the design of PKS modules. Although considerable optimization may be needed in adapting modules to unnatural substrates or in developing effective combinatorial biosynthesis approaches, the detailed insights revealed in this study provide a new structural framework from which to pursue these efforts.

METHODS SUMMARY

PikAIII for cryo-EM analysis was purified by Ni-affinity chromatography followed by duplicate gel filtration steps. To trap PikAIII in different states, the purified proteins were incubated with substrate, and adsorbed on holey carbon grids followed by blotting and vitrification. Bottom-up liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC/FT-ICR MS) confirmed substrate loading onto the PikAIII variants. All specimens were imaged on a Tecnai F20 transmission electron microscope (FEI) equipped with a field emission gun operated at 120 kV. Images were recorded at a magnification of ×66,964 on a Gatan US4000 CCD camera and defocus values ranging from −1.5 to −3.5 μm. EMAN routines were used for particle picking, generation of an initial model, and model refinement. Crystal structures of domains were independently fit in the EM maps as rigid bodies. Functional assays that require intermediate processing by PikAIII detected TE-offloaded products from PikAIII–TE or PikAIII/PikAIV[15,16].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 May 2013; accepted 28 April 2014.
Published online 18 June 2014.

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75,** 311–335 (2012).
2. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* http://dx.doi.org/10.1038/nature13423 (this issue).
3. Kittendorf, J. D. & Sherman, D. H. Developing tools for engineering hybrid polyketide synthetic pathways. *Curr. Opin. Biotechnol.* **17,** 597–605 (2006).
4. Walsh, C. T. Combinatorial biosynthesis of antibiotics: challenges and opportunities. *ChemBioChem* **3,** 124–134 (2002).
5. Menzella, H. G. & Reeves, C. D. Combinatorial biosynthesis for drug development. *Curr. Opin. Microbiol.* **10,** 238–245 (2007).
6. Menzella, H. G., Carney, J. R. & Santi, D. V. Rational design and assembly of synthetic trimodular polyketide synthases. *Chem. Biol.* **14,** 143–151 (2007).

©2014 Macmillan Publishers Limited. All rights reserved

7. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nature Biotechnol.* **23,** 1171–1176 (2005).
8. Sherman, D. H. The Lego-ization of polyketide biosynthesis. *Nature Biotechnol.* **23,** 1083–1084 (2005).
9. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106,** 3468–3496 (2006).
10. Xue, Y., Zhao, L., Liu, H. W. & Sherman, D. H. A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proc. Natl Acad. Sci. USA* **95,** 12111–12116 (1998).
11. Hansen, D. A. *et al.* Biocatalytic synthesis of pikromycin, methymycin, neomethymycin, novamethymycin, and ketomethymycin. *J. Am. Chem. Soc.* **135,** 11232–11238 (2013).
12. Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl Acad. Sci. USA* **103,** 11124–11129 (2006).
13. Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. *Structure* **14,** 737–748 (2006).
14. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16,** 2093–2107 (2007).
15. Aldrich, C. C., Beck, B. J., Fecik, R. A. & Sherman, D. H. Biochemical investigation of pikromycin biosynthesis employing native penta- and hexaketide chain elongation intermediates. *J. Am. Chem. Soc.* **127,** 8441–8452 (2005).
16. Aldrich, C. C., Venkatraman, L., Sherman, D. H. & Fecik, R. A. Chemoenzymatic synthesis of the polyketide macrolactone 10-deoxymethynolide. *J. Am. Chem. Soc.* **127,** 8910–8911 (2005).
17. Wong, F. T., Chen, A. Y., Cane, D. E. & Khosla, C. Protein–protein recognition between acyltransferases and acyl carrier proteins in multimodular polyketide synthases. *Biochemistry* **49,** 95–102 (2010).
18. Caffrey, P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBioChem* **4,** 654–657 (2003).
19. Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14,** 898–908 (2007).
20. Bonnett, S. A. *et al.* Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a *cis*-alkene. *Chem. Biol.* **20,** 772–783 (2013).
21. Lyon, A. M., Dutta, S., Boguth, C. A., Skiniotis, G. & Tesmer, J. J. Full-length $G\alpha_q$–phospholipase C-β3 structure reveals interfaces of the C-terminal coiled-coil domain. *Nature Struct. Mol. Biol.* **20,** 355–362 (2013).
22. Strunk, B. S. *et al.* Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science* **333,** 1449–1453 (2011).
23. Nguyen, C. *et al.* Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505,** 427–431 (2014).
24. Masoudi, A., Raetz, C. R., Zhou, P. & Pemble, C. W. IV Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature* **505,** 422–426 (2014).
25. Bunkoczi, G. *et al.* Mechanism and substrate recognition of human holo ACP synthase. *Chem. Biol.* **14,** 1243–1253 (2007).

Acknowledgements This work was supported by the Pew Scholar Program in Biomedical Sciences (G.S.), the University of Michigan Biological Sciences Scholars Program (G.S.), Rackham Merit and American Foundation for Pharmaceutical Education predoctoral fellowships (D.A.H.), a National Research Service Award postdoctoral fellowship (J.A.C.), the Life Sciences Research Foundation (A.R.H.N.), National Institutes of Health grant 1R21CA138331-01A1 (K.H.), GM076477 (D.H.S. and J.L.S.), DK042303 (J.L.S.), and DK090165 (G.S.), and the Hans W. Vahlteich Professorship (to D.H.S.).

Author Contributions J.R.W. produced PikAIII variants and conducted enzyme assays. S.D. carried out cryo-EM data collection and processing. A.M.D. assisted with cryo-EM image processing. W.A.H., A.R.H.N. and K.H. carried out mass-spectrometry analysis. D.A.H. synthesized the thiophenol-activated pentaketide substrate. J.A.C. produced initial PikAIII samples. S.D., J.R.W., K.H., J.L.S. and G.S. analysed the data and interpreted results. D.H.S., J.L.S. and G.S. designed research. J.R.W., S.D., K.H., D.H.S., J.L.S. and G.S. wrote the manuscript.

Author Information The cryo-EM maps have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-5663 (pentaketide–KS_5–PikAIII), EMD-5664 (β-ketohexaketide–PikAIII), EMD-5665 (β-hydroxyhexaketide–PikAIII conformation i), EMD-5666 (β-hydroxyhexaketide–PikAIII conformation ii) and EMD-5667 (β-hydroxyhexaketide–PikAIII conformation iii). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.S. (skinioti@umich.edu), J.L.S. (JanetSmith@umich.edu), D.H.S. (davidhs@umich.edu) or K.H. (kicki@umich.edu).

©2014 Macmillan Publishers Limited. All rights reserved

TURNING POINT Christian scientist spreads message about climate change **p.567**

RECRUITMENT Swedish research grant leads to hiring drive for global scientists **p.567**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**



PURESTOCK/GETTY

SYMPOSIA

Behind the scenes

Early-career researchers who help to organize conferences develop crucial skills that go beyond just booking speakers.

BY CAMERON WALKER

When Erin Eastwood's adviser suggested in January that their laboratory should organize a symposium, the graduate student had several questions, including a very basic one: what exactly is a symposium? Now Eastwood, who is studying for a master's degree in conservation biology at Columbia University in New York, intimately understands the meaning of the word: along with several others in her lab, she has developed a symposium topic and talks, and has contacted researchers in the field, none of whom she had met before, to speak at the Society for Conservation Biology's Oceania Section meeting in Fiji in July.

As a result, Eastwood has considerably expanded her network in the South Pacific, where she hopes to continue doing research, and will soon meet her new contacts face to face. "This is great for me, to not only identify the people who I should be connecting with, but also getting to know them in person," she says.

Building a professional network by finding speakers is just one of the many benefits that early-career researchers can gain from organizing scientific meetings. Like Eastwood, they might also need to write abstracts for symposium topics and for individual talks. But the process might be even more involved — it could mean planning a year or more ahead of the event to find venues, apply for funding and approach conference sponsors. At the meeting, volunteers can hone their public-speaking skills by introducing speakers and kicking off follow-up questions after the talk. And they will find out how to develop a symposium that will address controversial issues in a balanced way, how to plan for a certain number of attendees, and how to determine the right length of time for a meeting to last. In addition, they will learn how to respond to unexpected challenges, such as last-minute programme changes or faulty audio-visual equipment.

Anne-Marie Glynn rapidly acquired novel skills as a PhD student while helping to organize the 2005 PhD Student Symposium for the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. She and other student organizers had to solicit funding from industry, where they had no contacts, and write financial and scientific reports for a funder, which they had never done before. Several speakers cancelled at the eleventh hour. In one case, she and other organizers arranged a remote-video presentation for a speaker in Australia who was no longer able to attend. "No matter how well prepared you are, you won't have all eventualities covered," she says, adding that the need to think quickly and to handle the unexpected makes conference organizing an excellent learning experience.

The annual EMBL PhD symposium is one of many conferences that are aimed at early-career scientists. PhD students are encouraged to organize the three-day event, helping them to learn about conference creating. And many other scientific organizations are also eager for young scientists to come on board during the planning of workshops and conferences for researchers at all levels.

Conferences are professional-development opportunities in terms of both the information shared there and the organizing process itself, says Jodi Wesemann, assistant director for educational research at the American Chemical Society in Washington DC. A conference is richer when its organizers are diverse (see 'Organizing for diversity'), and Wesemann says that early-career scientists are a key part of the scientific enterprise. She notes that such scientists often gravitate towards organizing symposia on cutting-edge topics and encourage more career- and professional-development sessions, as the organization's Younger Chemists Committee has done at past conferences.

© 2014 Macmillan Publishers Limited. All rights reserved

Less-experienced volunteers might fear that they do not have the skills or connections to organize a symposium. But Wesemann says that volunteering is a safe place to hone existing skills and develop new ones. First-time organizers can seek advice on new tasks from more-experienced colleagues in addition to enhancing their existing skills, such as improving fluency with abstract-submission software.

"A lot of symposium-organizing is making it happen," says Wesemann, and that could involve simply asking to be included.

As a graduate student in the Scott Polar Research Institute at the University of Cambridge, UK, Allen Pope volunteered as an organizer of a career-skills workshop sponsored by the UK Polar Network, a branch of the Association of Polar Early Career Scientists. With that and subsequent work for these organizations under his belt, he approached the cryosphere division of the American Geophysical Union (AGU), which focuses on Earth's frozen regions, to ask whether early-career scientists could help to put sessions together. He ended up co-convening a session on mountain glaciers and their response to climate change at the AGU Fall Meeting in 2011.

As a first-timer, Pope learned the nuts and bolts of operating electronic session timers and using audio-visual equipment so that he could help speakers who were unfamiliar with the technology — and also so that he knew how to use them when his own talk came up. It was "much less daunting" when he knew he did not have to fumble with the equipment, says Pope, now a postdoctoral researcher at the National Snow & Ice Data Center in Boulder, Colorado.

Praveena Chackrapani, a recent graduate of the master's programme in environmental monitoring and management at the Malaysia campus of the University of Nottingham in Semenyih, has become immersed in conference organizing. As a full-time coordinator for the Society for Conservation Biology's biennial Asia Section meeting in August, her tasks include sorting incoming abstracts and registrations; seeking sponsorships to offset conference and speaker costs; coordinating workshops held before and after the conference; planning event food menus with caterers; and working with a web manager to update and maintain the conference website.



ERIN EASTWOOD

"This is great for me, to not only identify the people who I should be connecting with, but also getting to know them in person."
Erin Eastwood

Chackrapani says that six months of conference organizing have provided an intensive dose of the soft skills that were tougher for her to pick up in her previous career as an engineer. And she has also widened her contact network to include people she has worked with locally and the hundreds of delegates and speakers who will arrive in Malaysia in a few weeks. "You have to speak to sponsors, you have to deal with government, you have to network with all these famous, world-renowned scientists," she says.

The organizing process itself can make this last step easier, giving a younger scientist the confidence to approach more-experienced and highly celebrated researchers. "A lot of graduate students and early-career faculty feel that they're not worthy of talking to the 'big' people," says Josh Drew, a conservation biologist at Columbia, and Eastwood's adviser. Connection to the conference or society gives them a way into the circle that forms, sometimes literally, around renowned scientists, he says. "This title — symposium organizer — gives [other scientists] a reason to speak with them. It gives that e-mail the gravitas of the organization behind it."

Junior scientists might feel unsure about requesting time away from their research to arrange conferences or workshops. But in many cases, such activities benefit the entire laboratory. By pulling together a symposium based on their lab's research interests, the students in Drew's Columbia laboratory not only took some work off his plate but also enhanced Drew's own network. "Students are able to identify collaborations that might not be on your radar as a faculty member," he says. Plus, he adds, being able to point to his current students' accomplishments in an arena outside the lab allows him to show potential graduate students what their career-development opportunities might be should they choose to join his laboratory.

Matt Mio is an organic chemist at the University of Detroit Mercy in Michigan and chairs an advisory board at the American Chemical Society that brings undergraduates and student organizers to meetings. He says that encouraging young scientists to get involved in conference organizing is important for senior researchers who are interested in training the next generation and keeping their fields alive, and that junior scientists need to develop skills beyond those taught in an academic setting, including leadership, communication, task delegation and people management. Organizing and managing conferences and their logistics is a way to develop and exercise these skills, he says. "Show me the other place in their training, not in the classroom or the lab, where they're getting that kind of instruction." ■

Cameron Walker *is a freelance writer in Santa Barbara, California.*

ORGANIZING FOR DIVERSITY

Balanced conferences encourage female scientists

When female scientists help to organize conferences, the speaker roster tends to shift in a way that may be good for other female researchers and, arguably, for science overall, according to a study published in January (A. Casadevall and J. Handelsman *mBio* **5,** e00846-13; 2014). The analysis, which included nearly 2,000 speakers at plenary sessions and symposia at several microbiology meetings between 2011 and 2013, showed that symposia organized by committees that included at least one woman had up to 74% more female speakers than symposia organized by all-male committees. Co-author Arturo Casadevall, a microbiologist and immunologist at the Albert Einstein College of Medicine at Yeshiva University in New York, said that although more data are needed, it is possible that female conveners are more aware of the need for gender balance and might seek out female speakers.

Some societies provide guidelines for achieving gender balance. For example, the European Molecular Biology Organization calls for at least 30% of the speakers on workshops and courses to be female. Other professional organizations have also begun to look at gender balance at conferences. The American Astronomical Society's Committee on the Status of Women in Astronomy has been tracking the numbers of male and female speakers at astronomy conferences and offers advice for organizing committees on how to increase diversity.

Katayun Kamdin, a physics graduate student at the University of California, Berkeley, used her role as a conference organizer to encourage women in her field. At the American Physical Society's West Coast Conference for Undergraduate Women in Physics in January, which drew more than 150 women, she and her colleagues brought in career panels that included female researchers in both industry and academic institutions. Post-conference surveys showed that attendees left more enthusiastic about their options, more confident about applying to graduate school and less likely to consider an academic track as their only choice. "It made a difference in what students could see as possible career paths," she says. C.W.

© 2014 Macmillan Publishers Limited. All rights reserved

TURNING POINT
Katharine Hayhoe

The US government last month released the third National Climate Assessment *— a report on climate-change impacts across the United States — with Canadian Katharine Hayhoe as a lead author. Hayhoe, who is director of the Texas Tech University Climate Science Center in Lubbock, describes her evolution as a science communicator.*



Did you set out to study climate change?
No. I double-majored in astronomy and physics at the University of Toronto in Canada, but my interests shifted after I took a course on climate change. I was shocked by the magnitude of the problem as well as by the fact that nonlinear fluid dynamics is at the core of climate models. All of a sudden, I realized I'd been developing the skills necessary to study an urgent global problem. After getting my master's degree at the University of Illinois at Urbana-Champaign, I decided to become a consultant to help industry, government and non-profit clients to assess the impacts of climate change.

Did any consulting job alter your career?
Yes, I started working on the impact of climate change on the ecology of the Great Lakes region with funding from the Union of Concerned Scientists in Cambridge, Massachusetts, and the Ecological Society of America in Washington DC. It was an eye-opener to find that ecologists were using 20-year-old climate projections, which were considered obsolete in atmospheric science; I realized how huge the gap was between climate science and all other areas of science. Working on the resulting report, *Confronting Climate Change in the Great Lakes Region*, was a pivotal moment for me because it clarified the importance of generating high-resolution information and translating it into something that ecologists and water managers could use. I became frustrated by the limitations of the available tools, so I decided to return to academia.

How did you end up in Texas?
While on sabbatical from Notre Dame University in Indiana, my husband, a linguistics professor and pastor, was invited to Lubbock to serve as an interim pastor at a local church. Texas Tech University had an atmospheric-sciences programme, so we decided to move, and I got a grant-supported research-professor position. Climate change isn't a popular topic in some US communities, especially in conservative states, so it was stressful to move to a place where we knew that at least some people in the community would be strongly opposed to what I do. But embracing the unexpected has been enormously positive for my career. I later went back to the University of Illinois to earn my PhD so I would be eligible for tenure at Texas Tech.

How did you become a voice for climate adaptation?
Soon after we moved to Texas, I was invited to speak at a women's group about climate change. The questions I got, such as "How do you know this isn't just a natural cycle?" and "Aren't the ice caps on Mars shrinking too, so it's the Sun's effect?", challenged me to find answers. Shortly thereafter, I started getting invitations to speak for other groups, and my husband's congregation began asking questions. Before long, we co-wrote a book, *A Climate for Change*. Writing this book felt like I was exposing my world views as a Christian. I believe the idea that scientists are completely objective is a myth. The only legitimate way to do science is to admit our views and, where appropriate, share them with others.

How did you approach the US *National Climate Assessment*?
We aimed to make sure the science could be communicated. We wanted to make the best-available science accessible to every person who is interested. So we included videos of scientists explaining climate change and answered common, specific questions, such as why we think it's human-caused instead of a part of a natural cycle. I think those efforts helped us to achieve the high level of interest that this report received compared to our past assessments. ■

INTERVIEW BY VIRGINIA GEWIN

FUNDING

Salk windfall

A US$25-million gift to the Salk Institute for Biological Studies in La Jolla, California, is creating hiring opportunities in four research areas for scientists at all career stages. The bequest will help the institute to move into innovative areas of research, says Marsha Chandler, Salk's executive vice-president. Salk is recruiting researchers in cancer, metabolism and neuroscience, and Chandler says that the gift will help the institute to bring in more computational infrastructure and expertise to support these and other research areas. The bequest is from philanthropist Conrad Prebys and puts Salk within sight of its $300-million fund-raising goal, Chandler says.

RECRUITMENT

Sweden looks abroad

The Karolinska Institute in Stockholm, Sweden, is seeking scientists in four human-health research areas after receiving 522 million kronor (US$78 million) from the Swedish Research Council towards international hiring. The institute has already recruited four top scientists: two from the United States, one from the University of Oxford, UK, and one from Singapore. It is hiring researchers at all career stages to work in four centres that will study eating disorders, cancer, stem cells and regenerative medicine, and psychiatric genomics. Hans-Gustaf Ljunggren, Karolinska's dean of research, says that the institute aims to expand and establish a stronger global presence. Birgitta Henriques-Normark, vice-dean for recruitment, is developing a strategic plan for hiring.

JOB SHARING

Game of clones

Four faculty members proposed taking job sharing to new levels this month when they jointly applied for the position of president–vice chancellor at the University of Alberta in Edmonton, Canada. Twelve other faculty groups followed suit, and in their applications the groups took a light-hearted jab at the growing economic and social divide between faculty and administration. Kathy Cawsey, associate professor of English at Dalhousie University in Halifax, led the effort. None of the groups has been contacted for a job interview.

© 2014 Macmillan Publishers Limited. All rights reserved



JACEY

A ONE-SIDED ARGUMENT

The voice of reason.

BY ALEX SHVARTSMAN

Our arguments became one-sided after the aliens crash-landed in Manhattan.

"It was an accident," Colette says. "A tragic mistake. They didn't mean to hurt anyone."

She prattles on and on like this. I sulk.

There may be 7 billion people on Earth, but our planet is mostly uninhabited, full of vast empty spaces, even if you only count the landmass. So the odds of those octopus-looking bastards crashing in one of the most densely populated cities in the world are, what — some fraction of a per cent? A dark cosmic joke, except that there are 20,000 dead people buried under the rubble, and their families aren't laughing.

Colette only sees the best in people, even when they aren't people at all. "They aren't perfect, but they aren't evil, either. It was a rough landing; their ship almost disintegrated on entry. It's not like they wanted to kill anyone."

Captain Sully managed to land a malfunctioning plane in the Hudson River some years back, and he was only human. These aliens, who built vessels capable of reaching another star, should have been able to do better. I bet they just didn't care.

"Years from now, that day won't be remembered for the tragedy and the loss, but as the moment we finally learned that we aren't alone in the Universe." Colette can keep up her side of the argument for hours. Days, even.

We're a real odd couple. I served two tours in Iraq, and she protested against the war. I'm quiet and she's chatty. But it was her non-stop chitchat that got me through the nightmares and the shakes and all the other fun parting gifts I returned home with after the war. PTSD, the doctors called it. I called it hell.

So I listen to her defend the aliens, and whenever I can no longer stand to hear it, I leave the house for a while.

I drive to Sunset Park, our favourite outdoor spot in Brooklyn from back when we were dating. We used to love coming here to take in the New York skyline. I sit at the bench where I proposed, and I stare at where the familiar contours of skyscrapers used to be. Manhattan still burns. Ash wafts across the water.

The alien ship levelled several city blocks downtown. The devastation they caused is worse than any war. Worse than 9/11. If some terrorist did this with a suitcase nuke, we'd be bombing their capital already. But these aliens, we just forgive them and welcome them with open arms? This isn't right.

I inhale the bitter smoke and it reminds me of death and pain and Fallujah. They can't get away with this. Somebody has to make the alien bastards pay.

Back home, I stare at the arsenal laid out on the workbench in my garage. The GLOCK 19 handgun and AR semi-automatic rifle have been there since the day of the crash. A pair of home-made IEDs have joined them since then. I've been busy.

I can get close. I can walk right up to the police barricades that cordon off the ship, and I can push my way past New York's Finest. They probably wouldn't try very hard to stop me; after dealing with the grieving families who show up there every day, the cops must hate the aliens nearly as much as I do. And when I get to the dozens of little octopuses repairing the hull of their ship, like they keep showing on the news, I can open fire. I bet I can take out a good number of them before anyone stops me.

The only thing holding me back is Colette.

She forgives them, utterly and without doubt.

I pace across the garage, and allow Colette's voice to soothe me.

Our arguments became one-sided after the aliens crash-landed in Manhattan. After their gargantuan spaceship crushed Colette's office building. They're one-sided because there's no use talking back to the dead.

All I can do instead is to honour her memory as best I can; to do what I've done every day since I lost her.

I take another longing look at the workbench and walk out of the garage. I go back to the house, back to our bedroom, where Colette's scent still lingers. I close my eyes, listen to her voice inside my head, and try very hard to let her persuade me. ■

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

Alex Shvartsman *is a writer and game designer from Brooklyn, New York. Read more of his fiction at www.alexshvartsman.com.*

© 2014 Macmillan Publishers Limited. All rights reserved